

EXPLORING THE GENETIC ARCHITECTURE AND IMPROVING GENOMIC
PREDICTION ACCURACY FOR YIELD, MINERAL CONCENTRATION, AND CANNING
QUALITY TRAITS IN COMMON BEAN (*PHASEOLUS VULGARIS*)

By

Paulo Cesar Izquierdo Romero

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Plant Breeding, Genetics and Biotechnology - Crop and Soils Sciences - Doctor of Philosophy

2023

ABSTRACT

Dry bean (*Phaseolus vulgaris* L.) is the most important legumes for human consumption worldwide and is an important source of protein, vitamins, and micronutrients in the human diet. This research aimed to *i*) uncover the genetic architecture of yield, Fe bioavailability and seed micronutrient concentration, *ii*) characterize the genetic control of canning quality traits, and *iii*) assess the accuracy of genomic prediction models for yield and end-use quality traits. The genetic architecture of yield and seed micronutrient concentration was assessed through a combination of meta-QTL analyses integrating published studies over the last two decades in dry bean. A Gaussian mixture model was used to determine the number of distinct QTL in the meta-QTL analyses. Consistent meta-QTL over different genetic backgrounds and environments were identified, reducing the confidence interval compared with initial QTL. Furthermore, a genome-wide association (GWA) study with 295 lines of the yellow bean collection and 82 yellow recombinant inbred lines identified a major QTL for Fe bioavailability related to the ground factor P gene. A black breeding panel with 415 lines was evaluated for yield and canning quality traits in two growing seasons. Consistent associations for color retention, appearance and texture of canned beans were identified across years. Genomic prediction models provided moderate to high accuracy for end-use quality traits on the yellow and black populations. The genomic prediction accuracy was related to the heritability of each trait, and improvement of accuracy was observed for complex traits when secondary traits were included in the model, while for traits with major QTL, the use of associated markers as fixed effects increased prediction ability. The use of meta-QTL analyses and GWA in this study lays a foundation of the genetic control of yield and end-use quality traits and reveals the potential of genomic prediction for these traits in dry beans.

Copyright by
PAULO CESAR IZQUIERDO ROMERO
2023

*Para lo más importante y hermoso de mi vida ...
mi Mapasita Hermosa*

TABLE OF CONTENTS

INTRODUCTION	1
BIBLIOGRAPHY	4
CHAPTER 1: META-QTL ANALYSIS OF SEED IRON AND ZINC CONCENTRATION AND CONTENT IN COMMON BEAN (PHASEOLUS VULGARIS L.)	6
BIBLIOGRAPHY	22
CHAPTER 2: COMBINATION OF META-ANALYSIS OF QTL AND GWAS TO UNCOVER THE GENETIC ARCHITECTURE OF SEED YIELD AND SEED YIELD COMPONENTS IN COMMON BEAN	28
BIBLIOGRAPHY	51
APPENDIX	59
CHAPTER 3: GWAS-ASSISTED AND MULTI-TRAIT GENOMIC PREDICTION FOR IMPROVEMENT OF SEED YIELD AND CANNING QUALITY TRAITS IN A BLACK BEAN BREEDING PANEL	70
BIBLIOGRAPHY	88
APPENDIX	92
CHAPTER 4: GENOME-WIDE ASSOCIATION AND GENOMIC PREDICTION FOR FE-ZN CONCENTRATION AND FE BIOAVAILABILITY IN A YELLOW BEAN COLLECTION OF DRY BEANS	104
BIBLIOGRAPHY	122
APPENDIX	127
CONCLUSIONS	136

INTRODUCTION

Phaseolus vulgaris is the most important grain legume among the twenty that are commonly consumed in the human diet (Joshi and Rao, 2017). Dry beans are an important source of nutrients in several African and Latin American countries (Blair et al., 2010) and due of their Fe and Zn concentration, common bean has been included as a target crop in biofortification programs for countries with widespread human nutritional deficiencies. Microelement deficiencies are among the most common and devastating global nutritional problems (Hirschi, 2009). Beans are produced on millions of hectares of land worldwide, and the United States is one of the top producers (FAO 2022). The dry bean production in the United States is concentrated in North Dakota, Michigan, Nebraska, and Minnesota. Dry bean is a diverse crop in terms of cultivation methods, types of environment, morphological variability and consumer preferences, which have determined its adaptation to different niches (Broughton et al., 2003). *P. vulgaris* originated in Mexico and was independently domesticated in Middle America and the Andes (Bitocchi et al., 2013). The genetic and geographic differences between both gene pools are represented by differences in seed size, seed nutritional quality, resistance to pathogens, yield, plant growing period and other morpho-agronomic traits (Pérez-Vega et al., 2010; Blair et al, 2010).

Dry bean is diploid ($2n = 22$) with a genome size of approximately 600 Mb, with a total of 27,433 loci and 9,562 alternatively spliced transcripts (Schmutz et al., 2014). Dry beans are mainly self-pollinating, which facilitates the generation of homozygous lines, and by that feature, a common breeding strategy in dry beans is single plant selection in early generations for plant adaptation, architecture, and seed type. In sum, dry bean breeding programs need a couple of years to select some candidate lines with high yield, high mineral concentration and good canning quality. Although plant breeding has produced gains in quantitatively inherited traits, genetic gains can be boosted with the use of genomic tools in dry beans.

A common method to identify loci associated with complex traits is QTL mapping, which is typically carried out in biparental populations. QTL studies are the first step to uncover the genetic control of quantitative traits. However, QTL mapping has two important limitations; first, only alleles that differ between the parents of the specific cross can be tested (Collard et al., 2005), and second, the relatively small number of recombinations that occur during the creation of the populations that is translated in a low mapping resolution.

Genome-wide association (GWA) in conjunction with biparental QTL mapping can help resolve some limitations of working within single crosses and small population sizes. GWA overcomes the allele limitation of the QTL analysis because it is possible to use diversity panels or natural populations and reduce the linkage disequilibrium due to the accumulation of recombination events that are captured in a more diverse panel of individuals (Korte and Farlow, 2013). GWA also has limitations, and the most important is linked to the population structure, but that limitation can be addressed at least in part with structure corrections as PCA and Kinship matrices (Soto-cerda and Cloutier, 2010). Combining biparental QTL analysis with GWA can pinpoint the genomic regions most relevant to use in plant breeding (Brachi et al., 2010).

The target of QTL and GWA analysis is to identify regions or candidate genes that are associated with relevant traits. In breeding programs, that information is used to design molecular markers and apply them in marker-assisted selection (MAS). There are some examples of MAS in common bean, in most cases the markers are associated with single genes or QTL for disease resistance. For example MAS have been applied successfully in the introgression of major genes that confer resistance to rust (De Souza et al., 2007), common bacterial blight (CBB) (Gilbertson et al., 2012), bean golden yellow mosaic virus (BGYMV), and bean common mosaic virus (BCMV) (Blair et al., 2007). Although there are some examples of MAS in quantitative traits such as white mold (Ender et al., 2008) and root rot resistant (Navarro et al., 2009), the use of MAS in quantitative traits is laborious and can increase the costs of selection due that would be necessary many molecular markers to select lines with superior performance. Genomic prediction (GP) can address this issue using molecular markers across whole-genome to estimate the effect of all loci and then predict the breeding values for target traits (Wang et al., 2015). GP is becoming more popular in plant breeding due to the drop in the price of sequencing and the improvement in statistical approaches that address the multicollinearity of molecular markers.

Although GP is relatively new in crop breeding, several studies have used this methodology to select complex traits including mineral concentration in wheat (Velu et al., 2016) and grain yield in maize (Beyene et al., 2015) with promising results. More applicable to bean genetic improvement, Spindel et al. (2016) proposed a new GP approach in rice, which they named GP + de novo GWA (GP+GWA). This new approach uses the results of GWA on the GP models as fixed variables, and it can increase the accuracy and lead to more efficient breeding programs to select genotypes in early generations or to pick the most promising parental lines to generate new

populations. The crucial step in the GP+GWA model is making a good choice of the variables to be fixed in the model (Spindel et al., 2016).

Objective

This study aims to uncover the genetic architecture of seed yield and end-use quality traits through a combination of QTL, GWAS and meta-QTL analysis, and assess the accuracy, and the effect of molecular markers as fixed variables in genomic prediction models for yield and end-use quality traits.

Dissertation outline

Chapter 1 is a meta-QTL study aimed at uncovering the genetic architecture of Fe and Zn levels in seeds. The study was performed using seven QTL populations that comprised intra- and inter-gene pool crosses. Consistent meta-QTL regions across populations and environmental and candidate genes were determined.

Chapter 2 is a meta-QTL analysis that describes a genome-wide landscape for the most consistent genomic regions associated with seed yield components through QTL and GWA identified over the last two decades in dry bean. The study was performed using published data from 21 independent studies reported under sufficient water and drought conditions. Meta-QTL were identified and can be used for developing markers for marker-assisted selection and into genomic selection models to assess their potential as fixed variables to increase prediction accuracies.

Chapter 3 assesses the accuracy of genomic prediction (GP) models and the use of markers identified through genome-wide association as fixed variables in GP. Across two growing seasons, a black breeding panel was evaluated for seed yield components and canning quality traits in Michigan. Significant molecular markers associated with these traits and strategies to increase accuracy of GP were identified.

Chapter 4 assesses the genetic variability of microelements concentration and Fe bioavailability and evaluated the accuracy of GP in these traits. Across 2 locations, a yellow diversity panel was evaluated for Fe-Zn concentration, Fe bioavailability and agronomic traits. The relationship between microelements concentration and Fe bioavailability was determined, significant SNPs were associated with these traits and strategies to increase prediction ability were identified.

BIBLIOGRAPHY

- Beyene, Y., Semagn, K., Mugo, S., Tarekegne, A., Babu, R., Meisel, B., Sehabiague, P., Makumbi, D., Magorokosho, C., Oikeh, S., Gakunga, J., Vargas, M., Olsen, M., Prasanna, B. M., Banziger, M., & Crossa, J. (2015). Genetic gains in grain yield through genomic selection in eight bi-parental maize populations under drought stress. *Crop Science*, 55(1), 154–163. <https://doi.org/10.2135/cropsci2014.07.0460>
- Bitocchi, E., Bellucci, E., Giardini, A., Rau, D., Rodriguez, M., Biagetti, E., Santilocchi, R., Spagnoletti Zeuli, P., Gioia, T., Logozzo, G., Attene, G., Nanni, L., & Papa, R. (2013). Molecular analysis of the parallel domestication of the common bean (*Phaseolus vulgaris*) in Mesoamerica and the Andes. *New Phytologist*, 197(1), 300–313. <https://doi.org/10.1111/j.1469-8137.2012.04377.x>
- Blair, M. W., Fregene, M. a, Beebe, S. E., & Ceballos, H. (2007). Marker-assisted selection in common beans and cassava. *Marker-Assisted Selection: Current Status and Future Perspectives in Crops, Livestock, Forestry and Fish*, 81–115. <https://doi.org/10.1007/s11032-007-9115-9>
- Blair, M. W., González, L. F., Kimani, P. M., & Butare, L. (2010). Genetic diversity, inter-gene pool introgression and nutritional quality of common beans (*Phaseolus vulgaris* L.) from Central Africa. *Theoretical and Applied Genetics*, 121(2), 237–248. <https://doi.org/10.1007/s00122-010-1305-x>
- Brachi, B., Faure, N., Horton, M., Flahauw, E., Vazquez, A., Nordborg, M., Bergelson, J., Cuguen, J., & Roux, F. (2010). Linkage and association mapping of *Arabidopsis thaliana* flowering time in nature. *PLoS Genetics*, 6(5), 40. <https://doi.org/10.1371/journal.pgen.1000940>
- Broughton, W. J., Hern, G., Blair, M., Beebe, S., Gepts, P., & Vanderleyden, J. (2003). Beans (*Phaseolus* spp.) – model food legumes. *Plant and Soil*, 252, 55–128.
- Collard, B. C. Y., Jahufer, M. Z. Z., Brouwer, J. B., & Pang, E. C. K. (2005). An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: The basic concepts. *Euphytica*, 142(1–2), 169–196. <https://doi.org/10.1007/s10681-005-1681-5>
- De Souza, T. L. P. O., Alzate-Marin, A. L., Dessaune, S. N., Nunes, E. S., De Queiroz, V. T., Moreira, M. A., & De Barros, E. G. (2007). Inheritance study and validation of SCAR molecular marker for rust resistance in common bean. *Crop Breeding and Applied Biotechnology*, 7(1), 11–15. <https://doi.org/10.12702/1984-7033.v07n01a02>
- Ender, M., Terpstra, K., & Kelly, J. D. (2008). Marker-assisted selection for white mold resistance in common bean. *Molecular Breeding*, 21(2), 149–157. <https://doi.org/10.1007/s11032-007-9115-9>
- Gilbertson, R. L., & Singh, S. P. (2012). Direct and marker-assisted selection for resistance to common bacterial blight in common bean. *Crop Science*, 52(4), 1511–1521. <https://doi.org/10.2135/cropsci2011.08.0445>

- Hirschi, K. D. (2009). Nutrient Biofortification of Food Crops. *Annual Review of Nutrition*, 29(1), 401–421. <https://doi.org/10.1146/annurev-nutr-080508-141143>
- Joshi, P. K., & Rao, P. P. (2017). Global pulses scenario: status and outlook. *Annals of the New York Academy of Sciences*, 1392(1), 6–17. <https://doi.org/10.1111/nyas.13298>
- Korte, A., & Farlow, A. (2013). The advantages and limitations of trait analysis with GWAS: a review. *Plant Methods*, 9(1), 29. <https://doi.org/10.1186/1746-4811-9-29>
- Navarro, F. M., Sass, M. E., & Nienhuis, J. (2009). Marker-facilitated selection for a major QTL associated with root rot resistance in snap bean (*Phaseolus vulgaris* L.). *Crop Science*, 49(3), 850–856. <https://doi.org/10.2135/cropsci2007.10.0570>
- Pérez-Vega, E., Pañeda, A., Rodríguez-Suárez, C., Campa, A., Giraldez, R., & Ferreira, J. J. (2010). Mapping of QTLs for morpho-agronomic and seed quality traits in a RIL population of common bean (*Phaseolus vulgaris* L.). *TAG. Theoretical and Applied Genetics. Theoretische Und Angewandte Genetik*, 120(7), 1367–1380. <https://doi.org/10.1007/s00122-010-1261-5>
- Schmutz, J., McClean, P. E., Mamidi, S., Wu, G. A., Cannon, S. B., Grimwood, J., Jenkins, J., Shu, S., Song, Q., Chavarro, C., Torres-Torres, M., Geffroy, V., Moghaddam, S. M., Gao, D., Abernathy, B., Barry, K., Blair, M., Brick, M. a, Chovatia, M., ... Jackson, S. a. (2014). A reference genome for common bean and genome-wide analysis of dual domestications. *Nature Genetics*, 46(November 2013), 707–713. <https://doi.org/10.1038/ng.3008>
- Soto-cerda, B. J., & Cloutier, S. (2010). Association Mapping in Plant Genomes. *Genetic Diversity in Plants*, 3, 29–54.
- Velu, G., Crossa, J., Singh, R. P., Hao, Y., Dreisigacker, S., Perez, P., & Arun, R. (2016). Genomic prediction for grain zinc and iron concentrations in spring wheat. *Theor Appl Genet*, 129, 1595–1605. <https://doi.org/10.1007/s00122-016-2726-y>
- Wang, X., Yang, Z., & Xu, C. (2015). A comparison of genomic selection methods for breeding value prediction. *Science Bulletin*, 60(10), 925–935. <https://doi.org/10.1007/s11434-015-0791-2>

CHAPTER 1: META-QTL ANALYSIS OF SEED IRON AND ZINC CONCENTRATION AND CONTENT IN COMMON BEAN (*PHASEOLUS VULGARIS* L.)

Reproduced with permission from Springer Nature

Source:

This chapter has been published in Theoretical and Applied Genetics: Izquierdo, P., Astudillo, C., Blair, M. W., Iqbal, A. M., Raatz, B., & Cichy, K. A. (2018). Meta-QTL analysis of seed iron and zinc concentration and content in common bean (*Phaseolus vulgaris* L.). 131(8), 1645–1658. Available at: <https://doi.org/10.1007/s00122-018-3104-8>

Abstract

Common bean (*Phaseolus vulgaris* L.) is the most important legume for human consumption worldwide and it is an important source of microelements, especially iron and zinc. Bean biofortification breeding programs develop new varieties with high levels of Fe and Zn targeted for countries with human micronutrient deficiencies. Biofortification efforts thus far have relied on phenotypic selection of raw seed mineral concentrations in advanced generations. While numerous quantitative trait loci (QTL) studies have been conducted to identify genomic regions associated with increased Fe and Zn concentration in seeds, these results have yet to be employed for marker-assisted breeding. The objective of this study was to conduct a meta-analysis from seven QTL studies in Andean and Middle American intra- and inter-gene pool populations to identify the regions in the genome that control the Fe and Zn levels in seeds. Two meta-QTL specific to Fe and two meta-QTL specific to Zn were identified. Additionally, eight Meta QTL that co-localized for Fe and Zn concentration and/or content were identified across seven chromosomes. The Fe and Zn shared meta-QTL could be useful candidates for marker-assisted breeding to simultaneously increase seed Fe and Zn. The physical positions for 12 individual meta-QTL were identified and within five of the meta-QTL, candidate genes were identified from six gene families that have been associated with transport of iron and zinc in plants.

Introduction

Common bean (*Phaseolus vulgaris* L.) is the most important grain legume among the twenty that are commonly consumed in human diets (Joshi and Rao 2017; McClean et al. 2004). Common bean is a diverse crop in terms of morphological and seed variability, cultivation methods, environmental adaptation and consumer preferences and these factors have made it suitable for many different niches (Broughton et al. 2003). *Phaseolus vulgaris* was reported to have been domesticated independently twice, in Mexico and the Andes (Peru, Colombia) (Bitocchi et al. 2013). Members of the two gene pools, Andean and Middle American, vary in seed size, seed nutritional quality, resistance to pathogens, yield, days to maturity, and other morpho- agronomic traits (Pérez-Vega et al. 2010; Blair et al. 2010a).

Dry bean is an important dietary source of iron and zinc (Bouis and Welch 2010; Carrasco-Castilla et al. 2012). It has been included as a target crop in biofortification programs for countries with widespread human nutritional deficiencies. Microelement deficiencies are among the most common and devastating global nutritional problems (Hirschi 2009). In terms of seed

micronutrient concentrations, bean germplasm from the Andean and Middle American gene pools show some variability, Blair et al. (2010a) reported that Andean gene pool and inter-gene pool crosses tend to have higher concentrations of minerals than Middle American beans. The natural variability for seed Fe and Zn has been utilized in breeding programs based on phenotypic selection for seed mineral concentrations (Blair et al. 2010b). However, phenotypic selections alone are not leading to the genetic gains needed to make biofortification successful (Vasconcelos et al. 2017). Genomic tools are needed to help breeders to reach target concentrations in dry beans.

One of the most common strategies to unravel quantitative traits are Quantitative Trait Loci (QTL) studies. Several QTL analyses have been conducted in common bean to identify regions associated with seed Fe and Zn levels (Guzman- Maldonado et al. 2003; Cichy et al. 2009; Blair et al. 2010d; Blair et al. 2011, 2013; Blair and Izquierdo 2012). The use of common markers across different maps makes it possible to integrate such QTL in order to improve the accuracy of positioning and decrease the confidence interval using meta-QTL analysis.

Meta-QTL analysis compiles information from multiple studies and improves QTL location by comparing individual experiments and narrowing down confidence intervals obtained from individual analyses (Goffinet and Gerber 2000). Meta-QTL analysis has been conducted in several crops for various traits, including grain size and resistance to African gall midge in rice (Daware et al. 2017; Wu et al. 2016; Yao et al. 2016), grain traits in wheat (Tyagi et al. 2015), oil and protein in soybean (Van and McHale 2017) and resistance to white mold in common bean (Vasconcellos et al. 2017). Various statistical methods have been developed for meta-QTL analysis. The software program Biomeqator uses the transformed Akaike classification criterion (AIC) to determine the real number of QTL in a specific region (Arcade et al. 2004). To date, only one meta-QTL study for seed Fe and Zn has been published. That study was conducted with five maize populations and it resulted in the discovery of ten meta-QTL involved in Fe and/or Zn accumulation. The phenotypic variation contributed to the 10 MQTL ranged from 9 to 28% (Jin et al. 2013). Genomic advances have led to the identification of several key nutrient-regulation-related genes relevant to biofortification (Carvalho and Vasconcelos 2013). The accumulation of Fe and Zn in seeds is determined through several mechanisms. Gene families involved in root mineral uptake include ZIP (Zinc/Iron-regulated transporter-related protein), ZIF (zinc-induced facilitator), HMA (heavy metal associated), FRO (ferric reductase oxidase), and NA (nicotianamine) (Haydon and Cobbett 2007; Curie et al. 2009; Haydon et al. 2012). Shoot transport

gene families include ZIP, FRO, NA and MATE (multidrug and toxic compound extrusion) (Vert et al. 2002; Wintz et al. 2003; Bashir et al. 2006; Ishimaru et al. 2005, 2011, 2012) and seed filling genes include HMA and NRAMP (Natural Resistance Associated Macrophage Protein) (Connorton et al. 2017a; Mary et al. 2015; Carvalho and Vasconcelos 2013; López-Millán et al. 2016). Genes in the NA, IRT (iron-regulated transporter), VIT (vacuolar iron transporter) and ferritin families have been used to increase the concentration of Fe and/or Zn in grain of wheat and rice (Borg et al. 2012; Moreno-Moyano et al. 2016; Connorton et al. 2017b; Singh et al. 2017).

In this study, individual genetic maps and QTL for seed Fe and Zn concentration and content from seven populations were used to develop a consensus map and to identify Fe and Zn meta-QTL. These meta-QTL narrowed down confidence intervals of initial individual analysis' and the physical regions of the meta-QTL were identified. Furthermore, candidate genes within the meta-QTL intervals that belong to families of genes that have been reported in the literature in the process of uptake, transport and/or remobilization of Fe and Zn were selected.

Materials and methods

Populations

QTL information related to Fe and Zn in common bean seed was collected for seven populations, five of which have been previously published. We have summarized the details in Table 1. The seven populations include two Andean (AND 696 × G19833 (AG), G21242 × G21078 (GG2)), two Middle American (G14519 × G4825 (GG1), Black Magic × Shiny Crow (BS)), and three inter-gene pool (DOR 364 × G19833 (DG), BAT 93 × Jalo EEP (BJ), Cerinza × G10022 (CG)) populations of common bean. Additional description of field trials, statistical analysis, and molecular markers for the published studies have been previously reported in detail (Blair et al. 2009; Cichy et al. 2009, Blair et al. 2010c, 2011, 2013; Blair and Izquierdo 2012). The description of field trials and the map development for BS has been previously reported (Cichy et al. 2014).

Phenotypic data

The seven populations were planted in multiple years, environments, and locations. The GG1 and GG2 populations were planted in three locations in Colombia: Darien, Palmira and Popayan. DG was planted in Popayan and Darien and CG in Palmira and Darien. The AG population was planted in Darien under high and low soil P treatments. The BS population was planted in Michigan, USA and the BJ population in Darien, Colombia (Table 2).

Table 1.1 A summary of the common bean mapping populations considered for meta-QTL analysis Markers.

ID	Parents	Genepool	Type of population	Population Size	No. of markers	Number of QTLs
AG	AND696 x G19833	A x A	RILs	77	167	23
GG2	G21242 x G21078	A x A	RILs	100	118	9
GG1	G14519 x G4825	M x M	RILs	110	114	17
BS	Black Magic x Shiny	M x M	RILs	92	681	NR
DG	DOR364 x G19833	M x A	RILs	87	499	26
BJ	BAT93 x Jalo EEP	M x A	RILs	78	458	NR
CG	Cerinza x G10022	A x W	AB	138	143	13

Gene pool: A andean, M middle American, W wild NR not previously reported Conc concentration AG: Cichy et al. (2009), GG2: Blair et al. (2011), GG1: Blair et al. (2010c), BS: Cichy et al. (2014), DG: Blair et al. (2009), Galeano et al. (2011), BJ: Freyre et al. (1998), CG: Blair and Izquierdo (2012), Blair et al. (2013).

Table 1.2 Summary of the location, years, range of Fe and Zn, and environmental conditions of QTL studies included for meta-QTL analysis.

Population	Location	Year	Range ppm		MASL	AT (°C)	AR (mm)	PH
			Fe	Zn				
AG	Darien - Colombia	2000	53-77	16-29	1485	20	1,288	5.6
	Darien - Colombia	2005	39-79	20-23				
GG2	Darien - Colombia	2011	28-95	17-37	1450	20	1,650	5.6-6.1
	Palmira - Colombia	2011	48-93	17-49	1000	24	905	7.8
	Popayan - Colombia	2011	30-88	22-57	1730	18	2,124	5.6-6.1
GG1	Darien - Colombia	2010	35-81	17-39	1400	20	1,500	5.6
	Palmira - Colombia	2010	43-97	17-32	1000	24	905	7.8
	Popayan - Colombia	2010	44-77	30-49	1730	18	2,124	6.1
BS	Richville, MI - US	2010	102-48	37-24	190	22	871 ^a	7.9
		2011	99-48	38-23				
DG	Popayan - Colombia	1998	40-79	19-37	1730	18	2,124	5.6
	Darien - Colombia	2003	42-84	17-42	1400	20	1,650	5.6
BJ	Darien - Colombia	2007	46-114	20 - 57	1450	20	1,650	5.6
CG	Palmira - Colombia	2012	54-100	27-39	996	24	950	7.2
	Darien - Colombia	2012	58-92	23-38	1485	20	1,288	5.6

MASL Meter above sea level, AT average yearly temperature in °C, AR average yearly rainfall.

^a Average rainfall between June and September.

Two methods of mineral analysis were implemented in the studies, Inductively Coupled Plasma–Optical Emission Spectrometry (ICP) and Atomic Absorption Spectroscopy (AAS) as described in (Blair et al. 2009). AAS was used to quantify minerals in the AG population and ICP was used in the BS population. The other studies used both methodologies AAS and ICP to

quantify Fe and Zn. All seven populations have data for seed Fe and Zn concentration. In addition, the CG and GG1 populations have data for seed Fe and Zn content (i.e., $\mu\text{g}/\text{seed}$), DG and CG have data for seed Fe and Zn concentration in cotyledon tissue, and CG also has seed coat Fe and Zn concentration data. The DG data on Fe and Zn concentration in cotyledon have not been published previously. Cotyledon concentration measurements were conducted as described by Blair et al. (2013) where 12 g of seed was washed with sterile water and peeled by hand using a sterile scalpel to remove the seed coat from the cotyledons. The cotyledon samples were dried before grinding. Dry cotyledon tissue powder was weighed into two replicates of ~ 0.25 g each and was analyzed separately for Fe and Zn concentration via AAS and ICP and the values of the replicates were averaged. Correlations between seed Fe- and Zn-related phenotypes were determined using with Pearson correlations using SAS software, v 9.3 (SAS-Institute 2011).

QTL analysis

The seven populations range in size from 77 lines in AG to 138 lines in CG. The number of markers ranges from 114 in GG1 to 681 in BS. Eighty-eight QTL for seed Fe and Zn concentration and/or content have been reported for the populations AG, CG, DG, GG1, and GG2 (Blair et al. 2009, 2010c, 2011, 2013; Cichy et al. 2009; Blair and Izquierdo 2012). While the DG QTL study reported previously found 26 QTL for seed Fe and Zn concentration (Blair et al. 2009), the QTL data we present here are based on a reanalysis of the phenotypic data with a saturated genetic map of this population that became available following the publication of the original DG QTL study (Galeano et al. 2011). This new map provided high resolution for QTL detection, and for that reason we repeated the QTL analysis with the new genetic map. We conducted a new QTL analysis for the BS and BJ populations via composite interval mapping (CIM). This is a new analysis that has not been previously published. The QTL analysis was carried out using the program Windows QTL cartographer version 2.5 (Wang et al. 2012). To identify an accurate significance threshold for each trait, an empirical threshold was determined using 1000 permutations (Churchill and Doerge 1994).

Map projection and meta-analysis

In total, data from 87 QTL were used for analysis (Table 1, Tables S1). In the population AG, all published QTL were used for meta-analysis except the QTL reported on chromosome 11. In AG chromosome 11 is composed of 15 random amplified polymorphic DNA (RAPD), 1 amplified

fragment length polymorphism (AFLP), and 1 simple sequence repeat (SSR). Due to the lack of sequence information for these markers, it was not possible to use these QTL in the meta-analysis.

The program Biomercator v3.0 (Arcade et al. 2004) was used to develop a consensus map and for subsequent meta-QTL analysis. The DG map was chosen as the anchor for the consensus map since it has a high marker density, including 499 molecular markers of which 462 are SSR or SNPs that have information about their physical position (Galeano et al. 2011). The other six populations were projected onto the DG map to integrate the seven populations into the consensus map. Six of the seven maps share common markers, the exception is that of the BS population which is composed of 681 SNPs (Cichy et al. 2014). The BS population was integrated in the analysis using the physical position of SSR and SNPs of the DG map in the *P. vulgaris* v.2.1 reference genome available in Phytozome v12 (Goodstein et al. 2012). The DG SSR sequences were obtained from Phaseolus genes website at <http://phaseolusgenes.bioinformatics.ucdavis.edu>. Using the DG SSRs and SNPs and BS SNPs physical positions, we made a bridge between both genetic maps, taking into account the distance in cM estimated by Schmutz et al. 2014. Positions of Fe and Zn concentration and content QTL were extrapolated onto the consensus map on the basis of common genetic marker positions. Co-localization of meta-QTL was determined by the Akaike's information criterion (AIC) (Akaike 1974), and the lowest value was considered the best fit model for meta-QTL prediction.

Candidate genes

The search for candidate genes was performed based on the physical positions of the meta-QTL regions. The most recent annotated version of the *P. vulgaris* reference genome v.2.1 in Phytozome was used to identify the physical position of the meta-QTL and genes contained in these regions (Goodstein et al. 2012). The candidate genes were selected on the basis of a literature review of genes that have been reported as having a role in root uptake, transport and accumulation of Fe and Zn in plants (Vert et al. 2002; Wintz et al. 2003; Ishimaru et al. 2005, 2011, 2012; Bashir et al. 2006; Haydon and Cobbett 2007; Curie et al. 2009; Borg et al. 2012; Haydon et al. 2012; Carvalho and Vasconcelos 2013; Mary et al. 2015; López-Millán et al. 2016; Moreno-Moyano et al. 2016; Connorton et al. 2017a, b; Singh et al. 2017). In the case of the bZIP family, this is a large gene family but only two genes of that family have been reported to be involved in plant mineral uptake and translocation in *Arabidopsis* (Assuncao et al. 2010); the

protein sequences of those two transcription factors were aligned to the *P. vulgaris* reference genome v.2.1 using BLASTP in NCBI (<https://blast.ncbi.nlm.nih.gov/>).

Results

Single population QTL analysis

This study combines 87 QTL for seed Fe and Zn concentration and content, including 39 for Fe and 48 for Zn across seven common bean populations. It includes 56 QTL collated from previous studies within the AG, CG, GG1, and GG2 populations (Table S2). Additionally, 31 previously unreported QTL were identified with CIM analysis in the BJ, BS and DG populations (Table 1). Among the QTL identified with CIM, three were identified for Fe seed cotyledon and one for Zn seed cotyledon concentration found in DG population. The QTL for cotyledon Fe are on chromosomes Pv02, Pv08 and Pv11, while the QTL for cotyledon Zn is on chromosome Pv11. The other 27 QTL of populations BJ, BS and DG include 11 QTL for seed Fe and 16 QTL for seed Zn concentration. The seed Fe/Zn concentration QTL are distributed on chromosomes Pv01, Pv02, Pv03, Pv04, Pv05, Pv06, Pv08, Pv09 and Pv11. R² values for Fe concentration QTL ranged from 0.08 on chromosome Pv03 to 0.27 on chromosome Pv11, and R² values for Zn concentration QTL ranged from 0.09 on chromosome Pv02 to 0.27 on chromosome Pv04 (Table S1).

Phenotypic values for seed Fe concentration ranged from 28 to 114 ppm and for seed zinc concentration from 16 to 57 ppm (Table 2). Seed Fe and Zn concentration were positively correlated across the seven populations and the average of those correlations was $r = 0.66$, 0.67 , and 0.50 for the Andean (AG, GG2), Middle American (GG1, BS) and inter-gene pool crosses (DG, BJ, CG), respectively. The average correlation between Fe and Zn concentration on the seven populations was 0.59 .

Chromosomes Pv01 and Pv06 contained the highest number of QTL while Pv10 was the only chromosome without a single QTL identified (Table S1). Individual QTL explained between 4 and 55% of the phenotypic variance. In total, 25 QTL had R² values greater than 20%, and, therefore, are considered major QTL. All QTL detected on Pv06 and Pv09 were detected only in intra-gene pool populations, whereas the remaining chromosomes (Pv01 through Pv05, Pv08, and Pv11) contained QTL from both intra- and inter-gene pool populations. Fifty-five percent of QTL had an Andean source while the other forty-five percent had Middle American sources. Pv02 had the most consistently detected QTL, such that QTL were detected in the same region in five out of the seven populations.

Consensus mapping

The DG genetic map was used as a reference to develop the consensus map. The DG map is highly saturated and, with the exception of the BS population, it had markers in common with all maps that were integrated into this study. The consensus map consists of 1038 markers with a total length of 2012 cM and an average distance between markers of 3.6 cM. Of the 87 QTL that were identified in the 7 populations, 72 were projected on the consensus genetic map (Table S3). The remaining 15 QTL that were not projected were in regions that did not have sufficient common markers to make a reliable projection on the consensus map. For consensus QTL projection, the chromosomal position, LOD score, and R² of the individual QTL were taken into consideration. Chromosome Pv06 had the highest number of consensus QTL (18 QTL) and Pv02 contained the highest number of QTL that came from different populations (5 populations) (Fig. 1). The order of each chromosome was estimated with the physical position of SSR and SNP markers in the *P. vulgaris* reference genome v.2.1.

Meta-QTL analysis

The meta-analysis of the 72 QTL projected in the consensus map was performed in Biomercator 3.0; the AIC was used to select the best QTL model on each chromosome (Table 3). The meta-analysis resulted in a genetic model with 12 meta-QTL that covered 47 of the 72 individual QTL from the seven populations (Fig. 2 and Table S4). The number of meta-QTL identified on each chromosome varied from one on chromosomes Pv01, Pv04, Pv09 and Pv11, and two on chromosomes Pv02, Pv06, Pv07 and Pv08. The mean R² of the MQTL ranged from 10.3 to 27.0%, while the 95% confidence intervals for the MQTL varied between 3.1 and 18.1 cM, with an average of 7.6 cM. The CI was narrower in all MQTL than the mean CI identified for the original QTL. The physical length of MQTL varied from 0.36 to 11.93 Mb. MQTL_Fe&Zn_6.1 and MQTL_Fe&Zn_6.2 contained the highest number of individual QTL, each one with 8 QTL. MQTL_Fe&Zn_11.1 contained QTL from three different inter-gene pool populations.

Three types of MQTL were identified: 1) Fe-MQTL, 2) Zn-MQTL, and 3) Fe & Zn-MQTL. Fe-MQTL included MQTL_Fe_7.2 and MQTL_Fe_8.1. MQTL_Fe_7.2 and MQTL_Fe_8.1 have two individual QTL each that come from two different populations, in the case of MQTL_Fe_7.2 the source in GG2 was the Andean line G21078 and for GG1 was the Middle America genotype G4825. Similarly, the sources of MQTL_Fe_8.1 came from both the Andean (Cerinza) and Middle

American (Black Magic) gene pools. Zn-MQTL are distributed on chromosomes Pv02 and Pv07. MQTL_Zn_2.2 includes QTL from two inter-gene pool populations (BJ and DG).

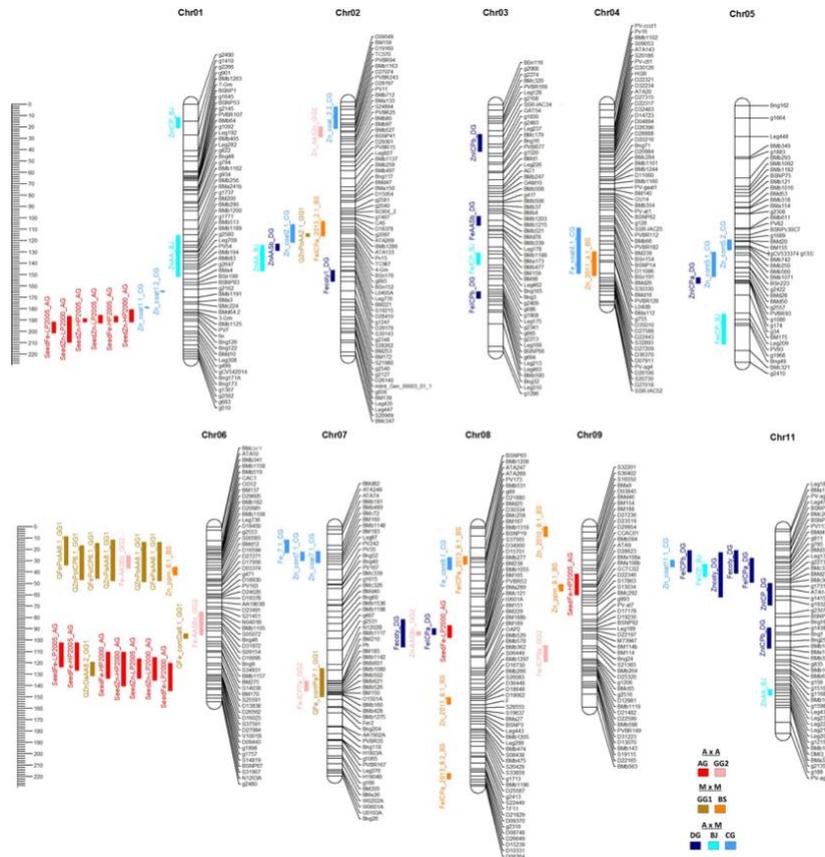


Figure 1.1 Mapping of QTL detected for iron and zinc, onto a single consensus map. AxA: Andean, MxM: Middle American, and AxM: inter-gene pool crosses.

The source of the BJ-QTL the source is the Middle American line BAT93 and in the DG-QTL the source is the Middle American DOR364. MQTL_Zn_7.1 is specific to CG and included two QTL that were identified in the CG cross between an Andean line and a Mexican wild genotype. One of these QTL is for seed Zn content with as source the wild genotype G10022, while the other QTL is for seed coat Zn content, with as source the Andean line Cerinza. The last group is the Fe & Zn-MQTL; this is the most abundant group with eight MQTL across seven chromosomes. MQTL_Fe&Zn_1.1 and MQTL_Fe&Zn_8.2 have Andean sources, MQTL_Fe&Zn_4.1 has Middle American sources and MQTL_Fe&Zn_2.1, MQTL_Fe&Zn_6.1, MQTL_Fe&Zn_6.2, MQTL_Fe&Zn_9.1 and MQTL_Fe&Zn_11.1 have both gene pools as sources of favorable alleles for Fe and Zn concentration/content in seed. The individual Fe and Zn contribution to the shared meta-QTL is available in Table S5.

Candidate genes

In total, 12 candidate genes were identified related to mineral transport or storage. These were found within 5 of the 12 MQTL regions (Table 4). The gene families identified in the MQTL have been previously reported to function in various points in Fe and Zn acquisition, including 1) Root uptake (ZIP, FRO and NA), 2) Translocation within the plant (ZIP, FRO, NA and MATE), and 3) Storage in seed (NRAMP). The name, position and family of each of the 15 candidate genes are reported in Table 4.

ZIP family: In both MQTL_Fe&Zn_9.1 and MQTL_Fe&Zn_11.1, there is a single ZIP family gene. Members of this family participate in mineral uptake, transport to leaves and translocation to seeds, embryo, endosperm, and seed coat (Vert et al. 2002; Ishimaru et al. 2005). The ZIP genes in MQTL_Fe&Zn_9.1 and MQTL_Fe&Zn_11.1 have both been annotated in the reference genome as zinc/iron trans- porters (Goodstein et al. 2012). In addition, a bZIP transcription factor was found in MQTL_Fe&Zn_11.1. bZIP transcription factors are associated with genes of the ZIP family and play a role in the uptake of minerals in plants (Assuncao et al. 2010). Although there are other bZIP elements in the MQTL, bZIP Phvul.011G035700 is the only one that aligned at the protein level with the transcription factors bZIP19 and bZIP23 (e value of 4×10^{-120} and 5×10^{-116} , respectively) that have an important function in Zn uptake capacity in Arabidopsis (Assuncao et al. 2010).

FRO family: Three FRO genes were identified within MQTL_Fe&Zn_6.1. FRO genes have important roles in iron uptake and in its transport in the vascular system (Wu et al. 2005; Kim and Guerinot 2007). Furthermore, other members of this family play an important role in chloroplast iron acquisition (Jeong et al. 2008).

NA family: MQTL_Fe&Zn_1.1 contains an NA family gene. NA genes have been related with mechanisms to acquire Fe and other minerals from the soil (Waters et al. 2006). NA chelates metal cations (Masuda et al. 2009) and there is evidence that suggests the NA family plays a role in the internal transport of Fe, Zn and other metals in plants (Takahashi et al. 2003; Schuler et al. 2012; Singh et al. 2017).

MATE family: Members of the MATE family are involved in the efflux of molecules from the cytoplasm to the outside of the cell or into the vacuole, and it is likely that these genes products export an Fe chelator that allows the movement of Fe in the plant (Grotz and Guerinot 2006;

Rogers et al. 2009). One MATE gene was found in MQTL_Fe&Zn_6.1, and two genes were in MQTL_Fe&Zn_4.1.

NRAMP family: NRAMP Genes are involved in transport of metals out of vacuoles. In Arabidopsis, members of this family are required for iron mobilization in germinating seeds (Thomine et al. 2003; Lanquar et al. 2010; Gollhofer et al. 2014; Mary et al. 2015). NRAMP genes were located in MQTL_Fe&Zn_1.1 and MQTL_Fe&Zn_9.1.

Table 1.3 Summary of the Meta-QTL associated with seed iron and zinc concentration.

MQTL ^a	Chr	Trait	Initial number of QTLs	Mean R2 (S. D.) ^b	Mean initial CI ^c (cM)	MQTL CI (cM)	CRL ^d	Physical position (Mb) ^e		MQTL Size (Mb)	No. genes in MQTL
								Start	End		
MQTL1.1	1	Fe - Zn	6	24.3 (9.3)	9.7	4.2	2.31	43.3	48.5	5.24	553
MQTL2.1	2	Fe - Zn	3	10.3 (5.69)	10.7	3.1	3.45	34.5	35	0.44	24
MQTL2.2	2	Zn	2	11.0 (2.83)	14.66	7.1	2.06	40.5	42.6	2.15	216
MQTL4.1	4	Fe - Zn	2	18.0 (12.73)	30.5	18.1	1.69	44.8	46	1.21	108
MQTL6.1	6	Fe - Zn	8	20.4 (5.07)	21.4	8.68	2.47	10.2	12.4	2.17	69
MQTL6.2	6	Fe - Zn	8	27.0 (11.3)	26.13	5.37	4.86	28.2	29.5	1.28	172
MQTL7.1	7	Zn	2	11.5 (2.12)	8.75	6.6	1.33	0.1	0.5	0.36	42
MQTL7.2	7	Fe	2	12.0 (1.41)	17	8.6	1.98	29.5	36.9	7.44	698
MQTL8.1	8	Fe	2	11.5 (0.71)	9.06	5.22	1.74	0.8	3.5	2.63	331
MQTL8.2	8	Fe - Zn	4	12.3 (3.77)	10.68	4.8	2.22	12.5	24.4	11.93	300
MQTL9.1	9	Fe - Zn	2	15.1 (5.66)	12	3.5	3.43	11.7	13.5	1.79	160
MQTL11.1	11	Fe - Zn	8	15.0 (7.27)	16.1	10.1	1.59	2.3	5.3	2.98	337

^a The Akaike's information criterion (AIC) was used to select the best QTL model for Fe and Zn, the one with the lowest AIC value was chosen as a significant model to indicate the number of MQTL on each chromosome. ^b S.D. = Standard deviation of R2 values of individuals QTL involved in MQTL. ^c CI = confidence interval of 95% of QTL location. ^d Coefficient of reduction in length from mean initial QTL to MQTL. ^e Physical size and position of MQTLs were estimated with the alignment of the SRAP sequences in the reference genome *P. vulgaris* v.2.1.

Discussion

Micronutrient deficiencies are widespread nutritional disorders affecting billions of people around the world (Nestel et al. 2006; Zhao et al. 2009; Vasconcellos et al. 2017). To date biofortification programs have increased Fe and Zn content in several crops; however, more efforts are still needed for at risk human populations to reach the recommended dietary requirements (Vasconcelos et al. 2017). Additional progress can be made through molecular breeding. Next-generation sequencing information has allowed genome sequencing of the most important crops to human consumption. There has also been progress in identifying genes that are

involved in the movement of Fe and Zn in plants and using these genes for biofortification of rice (Goto et al. 1999), cassava (Ihemere 2012), wheat (Borg et al. 2012), maize (Kanobe et al. 2013), lettuce (Goto et al. 2000) and soybean (Vasconcelos et al. 2014).

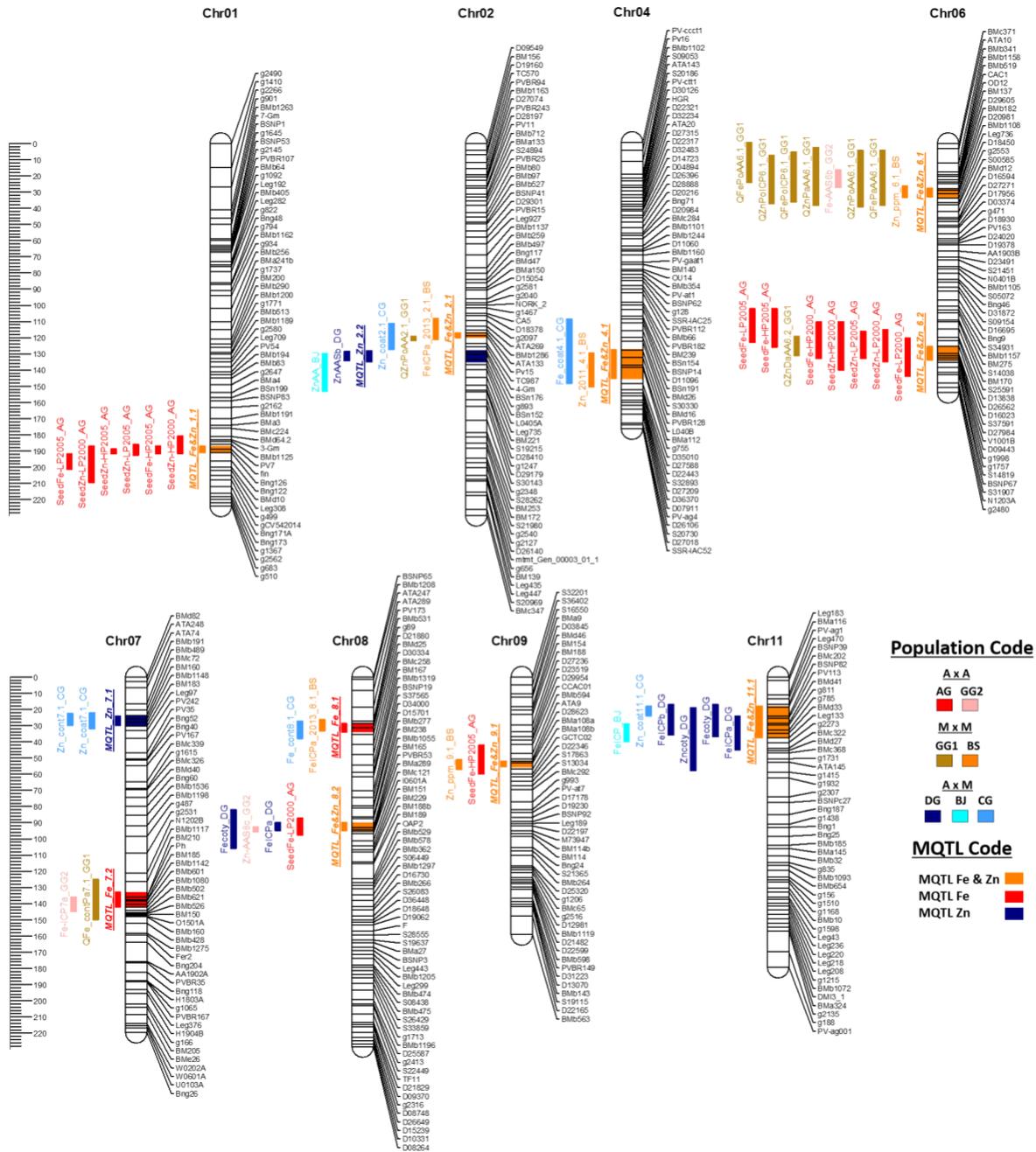


Figure 1.2 Meta-QTL analysis results for iron and zinc concentration from seven *P. vulgaris* QTL studies. AxM: Andean, MxM: Middle American, and AxM: inter-gene pool crosses.

Table 1.4 Candidate genes found within meta-QTL for seed Fe and Zn concentration and content.

MQTL	Candidate Gene Name	Chr	Start (bp)^a	End (bp)	Gene family
MQTL1.1	Phvul.001G177500	Chr01	43,453,927	43,463,444	NRAMP
MQTL1.1	Phvul.001G225000	Chr01	47,980,872	47,982,640	NA
MQTL4.1	Phvul.004G152100	Chr04	45,482,068	45,486,874	MATE
MQTL4.1	Phvul.004G152400	Chr04	45,504,102	45,509,691	MATE
MQTL6.1	Phvul.006G028701	Chr06	10,873,033	10,876,474	MATE
MQTL6.1	Phvul.006G030500	Chr06	11,701,246	11,702,422	FRO
MQTL6.1	Phvul.006G030550	Chr06	11,747,123	11,748,871	FRO
MQTL6.1	Phvul.006G030600	Chr06	11,804,817	11,806,566	FRO
MQTL9.1	Phvul.009G069700	Chr09	12,141,306	12,145,376	NRAMP
MQTL9.1	Phvul.009G077700	Chr09	13,064,171	13,066,165	ZIP
MQTL11.1	Phvul.011G035700	Chr11	3,293,008	3,295,446	bZIP
MQTL11.1	Phvul.011G058500	Chr11	5,228,604	5,232,591	ZIP

^a Physical position in the reference genome *P. vulgaris* v.2.1.

Seed Fe and Zn concentrations are quantitative traits with wide genotypic variability. As was reported by Beebe et al. (2001) and Blair et al. (2008), there is a difference in Fe and Zn concentration between the gene pools. Andean genotypes tend to have higher Fe but lower Zn than genotypes from the Middle American gene pool. Another important factor in the analysis of seed Fe and Zn concentration in common bean is the genotype–environmental interaction (GxE) (De Araújo et al. 2003; Pereira et al. 2014). Beebe (2012) and Hossain et al. (2013) reported that environmental factors such as soil characteristics and precipitation have an important influence in mineral accumulation in common bean.

To unravel the genetic complexity of seed Fe and Zn concentration and content, we collected data from seven populations over four locations, 9 years, and with multiple methodologies being used for mineral quantification, including whole seed, cotyledon, and seed coat measurements. The seed mineral data from the seven populations were positively correlated among locations, years and traits (Fe–Zn). The average correlation between Fe and Zn concentration in the seven populations was 59%. The correlation between Fe and Zn supports the well-reported observation that these traits are linked and if we increase the concentration of one of them, we will increase the other as well (Blair et al. 2010b; Blair and Izquierdo 2012). The correlation may be related to the similar movement of Fe and Zn through the plant, ultimately to the seed. Many of the same

genes are involved in both Fe and Zn transport (Kim and Guerinot 2007; Bashir et al. 2013). The positive correlation between seed Fe and Zn concentrations has been reported in other crops as well, including chickpea (Diapari et al. 2014; Upadhyaya et al. 2016).

In this study, we used seven populations that involve the two major gene pools of common bean. We included four intra- and three inter-gene pool crosses. In total, we tested 87 individual QTL detected in seven populations and were able to project 72 in the consensus map (41 have an Andean source, 25 have a Middle American source, and 6 have a wild Middle American source). The 72 QTL projected were distributed across all chromosomes except Pv10. The numerous QTL reflect the genetic complexity of the accumulation of Fe and Zn in common bean seeds.

The consensus map generated from the seven maps has a size of 2012 cM with 12 MQTL including two MQTL for Fe, two for Zn and 8 MQTL co-localized for Fe and Zn. It is interesting that of the nine QTL projected in the consensus map from the BS population, all but two clustered with QTL of other populations. All populations used in this study were planted in Colombia except the BS that was planted in Richville, MI–US, and although the environmental conditions are different (e.g., soil type, PH, average yearly rainfall), the BS-QTL are mainly close to the QTL of the other populations that have a Middle American genotype as a source. The proximity of BS-QTL with the Middle America QTL suggests that although there is a GxE interaction (De Araújo et al. 2003; Beebe 2012; Hossain et al. 2013 and Pereira et al. 2014) in the accumulation of Fe and Zn, the gene pool origin has an important effect in the accumulation of these minerals in the seed of common bean.

In this study, we narrowed down the CI in all 12 MQTL regions that allowed us to identify 12 candidate genes that could be responsible for some of the differences in the seed Fe and Zn concentration/content in the populations included in this study. Out of the 12 MQTL, the 8 Fe–Zn shared MQTL distributed over chromosomes 1, 2, 4, 6, 8, 9 and 11 have major potential for molecular breeding because they are associated with both Fe and Zn concentration and/or content and could potentially be used to increase the content of both elements in common bean seed. In five of these eight meta-QTL, there are 12 candidates for validation and subsequent application of allelic variation in breeding, e.g., by use of genetic transformation or allele screening in germplasm collections (ecoTilling). The genes Phvul.006G030500, Phvul.006G030550 and Phvul.006G030600 of the FRO family in MQTL_Fe&Zn_6.1 are of special interest, because genes of this family have been used successfully to increase mineral concentration in rice, wheat, and

soybean (Goto et al. 1999; Borg et al. 2012; Vasconcelos et al. 2014). FRO genes are responsible for reducing iron at the root surface (Wu et al. 2005; Mukherjee et al. 2006). Dicots acidify the rhizosphere to acquire Fe from the soil. The roots release organic acids and phenolic compounds to increase Fe³⁺ concentrations in the soil solution. These compounds chelate Fe³⁺ which subsequently is reduced to Fe²⁺ in the plasma membrane of root epidermal cells by ferric reductases which are encoded by members of the FRO gene family (Wu et al. 2005; Mukherjee et al. 2006; Connolly and Guerinot 2002; Kobayashi and Nishizawa 2012). We identified twelve candidate genes with the information available in the literature of the better-known gene families that have a relationship in the process to uptake, transport, and accumulation of Fe and Zn in plants. For the above, it is possible that we missed reporting some genes that belong to families that do not have a well-reported role in the movement of minerals in plants or genes that belong to families with unknown function.

Quantitative traits are a challenge in plant breeding due to the genetic complexity that governs these traits and the difficulty of stacking numerous alleles that control them. Meta-QTL analysis made possible the consolidation of 47 single QTL into 12 meta-QTL. These results showed a greater consolidation than a maize meta-QTL analysis for grain Fe and Zn, where 28 single QTL were consolidated into 10 meta-QTL (Jin et al. 2013). While we show that there are at least 12 regions that control the seed concentration/content of Fe and Zn in the common bean genome, the eight regions that associate with both Fe and Zn are most promising for focus in future studies. The stacking of eight independent regions in a single breeding line is challenging as there is a probability of one in 256 to stack the eight regions with the favorable alleles. The MQTL identified in this study have three potential uses, the first one is to generate markers for marker-assisted selection (MAS) for gene stacking breeding lines, the second way is the validation and subsequent use of the genes identified in this study in bean genetic transformation or eco-Tilling programs, and the last and perhaps most promising is the use of the MQTL regions in Genomic Selection (GS) models to increase the models accuracy in their use in bean breeding programs. Although this is a new study field, Spindel et al. (2016) reported promising results in rice using the regions identified by genome-wide association studies (GWAS) as fixed effects in GS models.

BIBLIOGRAPHY

- Akaike H (1974) A new look at the statistical model identification. *IEEE Trans Autom Control* 19:716–723
- Arcade A, Labourdette A, Falque M, Mangin B, Chardon F, Charcosset A, Joets J (2004) BioMercator: integrating genetic maps and QTL towards discovery of candidate genes. *Bioinformatics* 20:2324–2326
- Assuncao AGL, Herrero E, Lin Y-F, Huettel B, Talukdar S, Smaczniak C, Immink RGH, van Eldik M, Fiers M, Schat H et al (2010) Arabidopsis thaliana transcription factors bZIP19 and bZIP23 regulate the adaptation to zinc deficiency. *Proc Natl Acad Sci* 107:10296–10301
- Bashir K, Inoue H, Nagasaka S, Takahashi M, Nakanishi H, Mori S, Nishizawa NK (2006) Cloning and characterization of deoxymugineic acid synthase genes from graminaceous plants. *J Biol Chem* 281:32395–32402
- Bashir K, Nozoye T, Ishimaru Y, Nakanishi H, Nishizawa NK (2013) Exploiting new tools for iron bio-fortification of rice. *Biotechnol Adv* 31:1624–1633
- Beebe S (2012) Common bean breeding in the tropics. *Plant Breed Rev* 36:357–426
- Beebe S, Rengifo J, Gaitan E, Duque MC, Tohme J (2001) Diversity and origin of andean landraces of common bean. *Crops* 862:854–862
- Bitocchi E, Bellucci E, Giardini A, Rau D, Rodriguez M, Biagetti E, Santilocchi R, Spagnoletti Zeuli P, Gioia T, Logozzo G et al (2013) Molecular analysis of the parallel domestication of the common bean (*Phaseolus vulgaris*) in Mesoamerica and the Andes. *New Phytol* 197:300–313
- Blair MW, Izquierdo P (2012) Use of the advanced backcross-QTL method to transfer seed mineral accumulation nutrition traits from wild to Andean cultivated common beans. *Theor Appl Genet* 125:1015–1031
- Blair MW, Buendía HF, Giraldo MC, Métais I, Peltier D (2008) Characterization of AT-rich microsatellites in common bean (*Phaseolus vulgaris* L.). *Theor Appl Genet* 118:91–103
- Blair MW, Astudillo C, Grusak MA, Graham R, Beebe SE (2009) Inheritance of seed iron and zinc concentrations in common bean (*Phaseolus vulgaris* L.). *Mol Breed* 23:197–207
- Blair MW, Sharon JBK, Carolina A, Chee-Ming L, Andrea F, Grusak MA (2010a) Variation and inheritance of iron reductase activity in the roots of common bean (*Phaseolus vulgaris* L.) and association with seed iron accumulation QTL. *BMC Plant Biol* 10:215
- Blair MW, González LF, Kimani PM, Butare L (2010b) Genetic diversity, inter-gene pool introgression and nutritional quality of common beans (*Phaseolus vulgaris* L.) from Central Africa. *Theor Appl Genet* 121:237–248
- Blair MW, Medina JI, Astudillo C, Rengifo J, Beebe SE, Machado G, Graham R (2010c) QTL for seed iron and zinc concentration and content in a Mesoamerican common bean (*Phaseolus vulgaris* L.) population. *Theor Appl Genet* 121:1059–1070

- Blair MW, Monserrate F, Beebe S, Restrepo J, Flores J (2010d) Registration of high mineral common bean germplasm lines NUA35 and NUA56 from the red-mottled seed class. *J Plant Regist* 4:55
- Blair MW, Astudillo C, Rengifo J, Beebe SE, Graham R (2011) QTL analyses for seed iron and zinc concentrations in an intra-genepool population of Andean common beans (*Phaseolus vulgaris* L.). *Theor Appl Genet* 122:511–521
- Blair MW, Izquierdo P, Astudillo C, Grusak MA (2013) A legume biofortification quandary: variability and genetic control of seed coat micronutrient accumulation in common beans. *Front Plant Sci* 4:275
- Borg S, Brinch-Pedersen H, Tauris B, Madsen LH, Darbani B, Noeparvar S, Holm PB (2012) Wheat ferritins: improving the iron content of the wheat grain. *J Cereal Sci* 56:204–213
- Bouis HE, Welch RM (2010) Biofortification—a sustainable agricultural strategy for reducing micronutrient malnutrition in the global south. *Crop Sci* 50:S-20–S-32
- Broughton WJ, Hern G, Blair M, Beebe S, Gepts P, Vanderleyden J (2003) Beans (*Phaseolus* spp.)—model food legumes. *Plant Soil* 252:55–128
- Carrasco-Castilla J, Hernández-Álvarez AJ, Jiménez-Martínez C, Jacinto-Hernández C, Alaiz M, Girón-Calle J, Vioque J, Dávila-Ortiz G (2012) Antioxidant and metal chelating activities of peptide fractions from phaseolin and bean protein hydrolysates. *Food Chem* 135:1789–1795
- Carvalho SMP, Vasconcelos MW (2013) Producing more with less: strategies and novel technologies for plant-based food biofortification. *Food Res Int* 54:961–971
- Churchill GA, Doerge RW (1994) Empirical threshold values for quantitative trait mapping. *Genetics* 138:963–971
- Cichy KA, Caldas GV, Snapp SS, Blair MW (2009) QTL analysis of seed iron, zinc, and phosphorus levels in an andean bean population. *Crop Sci* 49:1742–1750
- Cichy KA, Fernandez A, Kilian A, Kelly JD, Galeano CH, Shaw S, Brick M, Hodgkinson D, Troxtell E (2014) QTL analysis of canning quality and color retention in black beans (*Phaseolus vulgaris* L.). *Mol Breed* 33:139–154
- Connolly EL, Guerinot M (2002) Iron stress in plants. *Genome Biol* 3:1024.1–1024.4
- Connorton JM, Balk J, Rodríguez-Celma J (2017a) Iron homeostasis in plants—a brief overview. *Metallomics* 9:813–823
- Connorton JM, Jones ER, Rodríguez-Ramiro I, Fairweather-Tait S, Uauy C, Balk J (2017b) Wheat vacuolar iron transporter TaVIT2 transports Fe and Mn and is effective for biofortification. *Plant Physiol* 174:2434–2444
- Curie C, Cassin G, Couch D, Divol F, Higuchi K, Le Jean M, Misson J, Schikora A, Czernic P, Mari S (2009) Metal movement within the plant: contribution of nicotianamine and yellow stripe 1-like transporters. *Ann Bot* 103:1–11

- Daware AV, Srivastava R, Singh AK, Parida SK, Tyagi AK (2017) Regional association analysis of MetaQTL delineates candidate grain size genes in rice. *Front Plant Sci.* <https://doi.org/10.3389/fpls.2017.00807>
- De Araújo R, Miglioranza É, Montalvan R, Destro D, Celeste M (2003) Genotype x environment interaction effects on the iron content of common bean grains. *Crop Breed Appl Biotechnol* 3:269–273
- Diapari A, Bett K, Deokar A, Warkentin TD, Tar'an BM (2014) Genetic diversity and association mapping of iron and zinc concentrations in chickpea (*Cicer arietinum* L.). *Genome* 57:459–468
- Freyre RO, Tsai SM, Gilbertson RL, Gepts P (1998) Towards an integrated linkage map of common bean 4. Development of an RFLP-based linkage map. *Theor Appl Genet* 85:513–520
- Galeano CH, Fernandez AC, Franco-Herrera N, Cichy KA, McClean PE, Vanderleyden J, Blair MW (2011) Saturation of an intra-gene pool linkage map: towards a unified consensus linkage map for fine mapping and synteny analysis in common bean. *PLoS ONE* 6:e28135
- Goffinet B, Gerber S (2000) Quantitative trait loci: a meta-analysis. *Genetics* 155:463–473
- Gollhofer J, Timofeev R, Lan P, Schmidt W, Buckhout TJ (2014) Vacuolar-iron-transporter1-like proteins mediate iron homeostasis in arabidopsis. *PLoS ONE* 9:1–8
- Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, Mitros T, Dirks W, Hellsten U, Putnam N et al (2012) Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res* 40:D1178–D1186
- Goto F, Yoshihara T, Shigemoto N, Toki S, Takaiwa F (1999) Iron fortification of rice seed by the soybean ferritin gene. *Nat Biotechnol* 17:282–286
- Goto F, Yoshihara T, Saiki H (2000) Iron accumulation and enhanced growth in transgenic lettuce plants expressing the iron-binding protein ferritin. *TAG Theor Appl Genet* 100:658–664
- Grotz N, Guerinot ML (2006) Molecular aspects of Cu, Fe and Zn homeostasis in plants. *Biochim Biophys Acta* 1763:595–608
- Guzman-Maldonado SH, Martinez O, Acosta-Gallegos JA, Guevara-Lara F, Paredes-Lopez O (2003) Putative quantitative trait loci for physical and chemical components of common bean. *Crop Sci* 43:1029–1035
- Haydon MJ, Cobbett CS (2007) A novel major facilitator superfamily protein at the tonoplast influences zinc tolerance and accumulation in arabidopsis. *Plant Physiol* 143:1705–1719
- Haydon MJ, Kawachi M, Wirtz M, Hillmer S, Hell R, Kramer U (2012) Vacuolar nicotianamine has critical and distinct roles under iron deficiency and for zinc sequestration in arabidopsis. *Plant Cell Online* 24:724–737
- Hirschi KD (2009) Nutrient biofortification of food crops. *Annu Rev Nutr* 29:401–421
- Hossain K, Islam N, Ghavami F, Tucker M, Kowalshy T et al (2013) Interdependence of genotype and growing site on seed mineral compositions in common bean. *Asian J Plant Sci* 12:11–20

- Ihemere U (2012) Iron biofortification and homeostasis in transgenic cassava roots expressing the algal iron assimilatory gene, FEA1. *Front Plant Sci* 3:1–22
- Ishimaru Y, Suzuki M, Kobayashi T, Takahashi M, Nakanishi H, Mori S, Nishizawa NK (2005) OsZIP4, a novel zinc-regulated zinc transporter in rice. *J Exp Bot* 56:3207–3214
- Ishimaru Y, Bashir K, Nishizawa NK (2011) Zn uptake and translocation in rice plants. *Rice* 4:21–27
- Ishimaru Y, Takahashi R, Bashir K, Shimo H, Senoura T, Sugimoto K, Ono K, Yano M, Ishikawa S, Arao T et al (2012) Characterizing the role of rice NRAMP5 in manganese, iron and cadmium transport. *Sci Rep* 2:1–8
- Jeong J, Cohu C, Kerkeb L, Pilon M, Connolly EL, Guerinot ML (2008) Chloroplast Fe(III) chelate reductase activity is essential for seedling viability under iron limiting conditions. *Proc Natl Acad Sci* 105:10619–10624
- Jin T, Zhou J, Chen J, Zhu L, Zhao Y, Huang Y (2013) The genetic architecture of zinc and iron content in maize grains as revealed by QTL mapping and meta-analysis. *Breed Sci* 63:317–324
- Joshi PK, Rao PP (2017) Global pulses scenario: status and outlook. *Ann N Y Acad Sci* 1392(1):6–17
- Kanobe MN, Rodermeier SR, Bailey T, Scott MP (2013) Changes in endogenous gene transcript and protein levels in maize plants expressing the soybean ferritin transgene. *Front Plant Sci* 4:1–14
- Kim SA, Guerinot ML (2007) Mining iron: iron uptake and transport in plants. *FEBS Lett* 581:2273–2280
- Kobayashi T, Nishizawa NK (2012) Iron uptake, translocation, and regulation in higher plants. *Annu Rev Plant Biol* 63:131–152
- Lanquar V, Ramos MS, Lelievre F, Barbier-Brygoo H, Krieger-Liszkay A, Kramer U, Thomine S (2010) Export of vacuolar manganese by AtNRAMP3 and AtNRAMP4 is required for optimal photosynthesis and growth under manganese deficiency. *Plant Physiol* 152:1986–1999
- López-Millán AF, Duy D, Philippar K (2016) Chloroplast iron transport proteins—function and impact on plant physiology. *Front Plant Sci* 7:1–12
- Mary V, Schnell Ramos M, Gillet C, Socha AL, Giraudat J, Agorio A, Merlot S, Clairet C, Kim SA, Punshon T et al (2015) Bypassing iron storage in endodermal vacuoles rescues the iron mobilization defect in the natural resistance associated-macrophage protein3 natural resistance associated-macrophage protein4 double mutant. *Plant Physiol* 169:748–759
- Masuda H, Usuda K, Kobayashi T, Ishimaru Y, Kakei Y, Takahashi M, Higuchi K, Nakanishi H, Mori S, Nishizawa NK (2009) Overexpression of the barley nicotianamine synthase gene HvNAS1 increases iron and zinc concentrations in rice grains. *Rice* 2:155–166
- McClellan P, Gepts P, Kami J (2004) Genomics and genetic diversity in common bean. In: Wilson RF, Stalker HT, Brummer EC (eds) *Legume crop genomics*. AOCS Press, Champaign, IL, pp 60–82

- Moreno-Moyano LT, Bonneau JP, Sánchez-Palacios JT, Tohme J, Johnson AAT (2016) Association of increased grain iron and zinc concentrations with agro-morphological traits of biofortified rice. *Front Plant Sci* 7:1–13
- Mukherjee I, Campbell NH, Ash JS, Connolly EL (2006) Expression profiling of the Arabidopsis ferric chelate reductase (FRO) gene family reveals differential regulation by iron and copper. *Planta* 223:1178–1190
- Nestel P, Bouis HE, Meenakshi JV, Pfeiffer W (2006) Symposium: food fortification in developing countries biofortification of staple food crops. *J Nutr* 136:1064–1067
- Pereira HS, Del Peloso MJ, Bassinello PZ, Guimarães CM (2014) Genetic variability for iron and zinc content in common bean lines and interaction with water availability. *Genet Mol Res* 13:6773–6785
- Pérez-Vega E, Pañeda A, Rodríguez-Suárez C, Campa A, Giraldez R, Ferreira JJ (2010) Mapping of QTL for morpho-agronomic and seed quality traits in a RIL population of common bean (*Phaseolus vulgaris* L.). *Theor Appl Genet* 120:1367–1380
- Rogers EE, Wu X, Stacey G, Nguyen HT (2009) Two MATE proteins play a role in iron efficiency in soybean. *J Plant Physiol* 166:1453–1459
- SAS-Institute (2011) SAS Institute Inc. 2011. SAS® 9.3 System Options: Reference, Second Edition
- Schmutz J, McClean PE, Mamidi S, Wu GA, Cannon SB, Grimwood J, Jenkins J, Shu S, Song Q, Chavarro C et al (2014) A reference genome for common bean and genome-wide analysis of dual domestications. *Nat Genet.* <https://doi.org/10.1038/ng.3008>
- Schuler M, Rellán-Álvarez R, Fink-Straube C, Abadía J, Bauer P (2012) Nicotianamine functions in the phloem-based transport of iron to sink organs, in pollen development and pollen tube growth in Arabidopsis. *Plant Cell* 24:2380–2400
- Singh SP, Keller B, Gruissem W, Bhullar NK (2017) Rice NICOTIANAMINE SYNTHASE 2 expression improves dietary iron and zinc levels in wheat. *Theor Appl Genet* 130:283–292
- Spindel JE, Begum H, Akdemir D, Collard B, Redoña E, Jannink J-L, McCouch S (2016) Genome-wide prediction models that incorporate de novo GWAS are a powerful new tool for tropical rice improvement. *Heredity (Edinb)* 116:395–408
- Takahashi M, Terada Y, Nakai I, Nakanishi H, Yoshimura E, Mori S, Nishizawa NK (2003) Role of nicotianamine in the intracellular delivery of metals and plant reproductive development. *Plant Cell* 15:1263–1280
- Thomine S, Lelièvre F, Debarbieux E, Schroeder JI, Barbier-Brygoo H (2003) AtNRAMP3, a multispecific vacuolar metal transporter involved in plant responses to iron deficiency. *Plant J* 34:685–695
- Tyagi S, Mir RR, Balyan HS, Gupta PK (2015) Interval mapping and meta-QTL analysis of grain traits in common wheat (*Triticum aestivum* L.). *Euphytica* 201:367–380
- Upadhyaya HD, Bajaj D, Das S, Kumar V, Gowda CLL, Sharma S, Tyagi AK, Parida SK (2016) Genetic dissection of seed-iron and zinc concentrations in chickpea. *Sci Rep* 6:24050

- Van K, McHale LK (2017) Meta-analyses of QTL associated with protein and oil contents and compositions in soybean [*Glycine max* (L.) Merr.] Seed. *Int J Mol Sci.* <https://doi.org/10.3390/ijms18061180>
- Vasconcellos RCC, Oraguzie OB, Soler A, Arkwazee H, Myers JR, Ferreira JJ, Song Q, McClean P, Miklas PN (2017) Meta- QTL for resistance to white mold in common bean. *PLoS ONE* 12:e0171685
- Vasconcelos MW, Clemente TE, Grusak MA (2014) Evaluation of constitutive iron reductase (AtFRO2) expression on mineral accumulation and distribution in soybean (*Glycine max*. L). *Front Plant Sci* 5:1–12
- Vasconcelos MW, Gruissem W, Bhullar NK (2017) Iron biofortification in the 21st century: setting realistic targets, overcoming obstacles, and new strategies for healthy nutrition. *Curr Opin Biotechnol* 44:8–15
- Vert G, Grotz N, Dédaldéchamp F, Gaymard F, Lou Guerinot M, Briat J-F, Curie C (2002) IRT1, an arabidopsis transporter essential for iron uptake from the soil and for plant growth. *Plant Cell* 14:1223–1233
- Wang S, Basten CJ, Zeng Z-B (2012) Windows QTL cartographer 2.5. Department of Statistics, North Carolina State University, Raleigh. <http://statgen.ncsu.edu/qtlcart/WQTLCart.htm>. Accessed 16 June 2017
- Waters BM, Chu H-H, DiDonato RJ, Roberts LA, Eisley RB, Lahner B, Salt DE, Walker EL (2006) Mutations in arabidopsis yellow stripe-like1 and yellow stripe-like3 reveal their roles in metal ion homeostasis and loading of metal ions in seeds. *Plant Physiol* 141:1446–1458
- Wintz H, Fox T, Wu YY, Feng V, Chen W, Chang HS, Zhu T, Vulpe C (2003) Expression profiles of arabidopsis thaliana in mineral deficiencies reveal novel transporters involved in metal homeostasis. *J Biol Chem* 278:47644–47653
- Wu H, Li L, Du J, Yuan Y, Cheng X, Ling HQ (2005) Molecular and biochemical characterization of the Fe(III) chelate reductase gene family in *Arabidopsis thaliana*. *Plant Cell Physiol* 46:1505–1514
- Wu Y, Huang M, Tao X, Guo T, Chen Z, Xiao W (2016) Quantitative trait loci identification and meta-analysis for rice panicle-related traits. *Mol Genet Genomics* 291:1927–1940
- Yao N, Lee C-R, Semagn K, Sow M, Nwilene F, Kolade O, Bocco R, Oyetunji O, Mitchell-Olds T, Ndjiondjop M-N et al (2016) QTL mapping in three rice populations uncovers major genomic regions associated with African rice gall midge resistance. *PLoS ONE* 11:e0160749
- Zhao FJ, Su YH, Dunham SJ, Rakszegi M, Bedo Z, McGrath SP, Shewry PR (2009) Variation in mineral micronutrient concentrations in grain of wheat lines of diverse origin. *J Cereal Sci* 49:290–295

**CHAPTER 2: COMBINATION OF META-ANALYSIS OF QTL AND GWAS TO
UNCOVER THE GENETIC ARCHITECTURE OF SEED YIELD AND SEED YIELD
COMPONENTS IN COMMON BEAN**

Abstract

Increasing seed yield in common bean could help to improve food security and reduce malnutrition globally due to the high nutritional quality of this crop. However, the complex genetic architecture and prevalent genotype by environment interactions for seed yield makes increasing genetic gains challenging. The aim of this study was to identify the most consistent genomic regions related with seed yield components reported in the last 20 years in common bean. A meta-analysis of QTL for seed yield components (MQTL-YC) was performed for 394 QTL reported in 21 independent studies under sufficient water and drought conditions. In total, 58 MQTL-YC over different genetic backgrounds and environments were identified, reducing three-fold on average the confidence interval (CI) compared with the CI for the initial QTL. Furthermore, 40 MQTL-YC identified were co-located with 210 SNP peak positions reported on GWAS, guiding the identification of candidate genes. Comparative genomics among of these MQTL-YC with MQTL-YC reported in soybean and pea allowed the identification of 14 orthologous MQTL-YC shared across species. The integration of MQTL-YC, GWAS and comparative genomics used in this study will be useful to refine and uncover the most consistent genomic regions related with seed yield components for their use in plant breeding.

Introduction

Common bean (*Phaseolus vulgaris* L.) is one of the most important legumes for direct human consumption around the world (Siddiq & Uebersax, 2022). Dry beans are a nutrient-dense food, rich in protein, fiber and micronutrients, and its consumption has been associated with health benefits in the prevention of cardiovascular disease, diabetes and obesity (Didinger et al., 2022). The ability for common bean to grow under low inputs, especially in low water and drought conditions along with the rich nutritional profile of dry bean, make it an appealing and a potential crop to face food security and climate change in the future (Medendorp et al., 2022; Siddiq and Uebersax, 2022). Additionally, legumes have up to 7-fold less greenhouse gas emissions compared with other crops such as wheat and canola (Jeuffroy et al., 2013), and improve the soil quality through nitrogen fixation and increasing soil carbon content (Jensen et al., 2012), which highlight the value of dry bean as a crop to improve agricultural and dietary sustainability.

Seed yields have incrementally increased in new dry bean cultivars over time and genetic gains have not plateaued (Vandemark et al., 2015). The use of new breeding tools could help continue that trend by more efficiently stacking positive alleles. Seed yield is a quantitative trait that is

controlled by numerous genes with small effects. Due to its complexity, seed yield has been divided into components, including number of pods per plant, seed size, and number of seeds per pod. Adams (1967) was the first to describe the interaction among these three yield components, where the increase in one of these three traits often resulted in a reduction of the others. Since dry bean has specific market class requirements for seed weight, changing this component is generally not a viable option to increase seed yield (Kelly, 2018). The total number of pods per plant and seeds per pod are more relevant for breeding within market classes, due to rigid seed size criteria.

Other traits associated with seed yield are related to dry matter partitioning toward seed, including pod harvest index and harvest index (Assefa et al., 2013; Polania et al., 2016; Nabateregga et al., 2019). These traits are important indicators of yield potential under drought and non-drought conditions. Under drought, the allocation of resources to reproductive growth is reduced, leading to flower and pod abortion in susceptible genotypes, while drought-tolerant genotypes continue the partition into the seed (Hageman and Volkenburgh, 2021). Phenology traits such as days to flowering and maturity are also associated with seed yield and may have different influence under drought or non-drought conditions. While longer days to maturity may be beneficial to increase seed yield under water sufficient conditions, under drought conditions short growing cycle minimize exposure to terminal drought, leading to better yield performance (Beebe et al., 2013; Vandemark et al., 2015; Keller et al., 2020). Environmental factors such as latitude, photoperiod, and temperature influence seed yield and component traits, and the identification of alleles for local vs broad adaptation will support genetic gains (MacQueen et al., 2021).

For the last 20 years, quantitative trait loci (QTL) and genome-wide association studies (GWAS) have been used to uncover the genetic architecture of seed yield components and hundreds of associated genomic regions have been identified. QTL analysis is a powerful tool to uncover the genetic architecture of complex traits. This technique comes with some limitations, including low allelic diversity and low recombination in mapping populations, which is reflected in a limited number of loci and recombination assessed on QTL analyses, resulting in QTL covering large genomic regions containing many genes (Brachi et al., 2010). GWAS includes diverse populations that ensure a higher allelic diversity than bi-parental populations and historic recombination events that overcome the two main limitations of QTL analysis (Korte and Farlow, 2013). GWAS has been reported to be a promising approach to identify quantitative trait nucleotides (QTN) associated with causative loci (Cano-Gamez and Trynka, 2020). Nevertheless,

most of GWAS are underpowered due to the limited population size and the small effect of causative loci in quantitative traits as seed yield (Evangelou and Ioannidis, 2013).

Although substantial efforts have been made to uncover the genetic architecture of seed yield components in dry bean through QTL and QTN this information is challenging to use in breeding due to the lack of standardized phenotyping and molecular markers, as well as genetic background and environmental effects that arise in multiple studies (Bernardo, 2008). Additionally, in species with more than one reference genome version, such as common bean, QTL and QTN physical position vary depending on the reference genome version used in each study. Since many market classes of dry beans are bred around the world the extrapolation of information generated among breeding programs may be limited (Vandemark et al., 2015). The Meta-QTL analysis (MQTL) is an approach that can overcome these limitations by integrating QTL from independent studies to identify and refine the most consistent QTL (Goffinet and Gerber, 2000; Izquierdo et al., 2018; Soriano and Alvaro, 2019), and the co-localization of QTN within MQTL regions leads to the identification of candidate loci with potential use in plant breeding (Shariatipour et al., 2021; Bilgrami et al., 2022).

The goals of this study were to *i*) perform a MQTL to uncover the genomic control of seed yield in dry bean *ii*) assess the co-localization of QTN within MQTL-YC *iii*) identify the physical position in the *P. vulgaris* v2.1 reference genome for all the QTL and QTN included in the analysis, and *iv*) evaluate the genomic collinearity of MQTL-YC regions of dry bean with MQTL and MGWAS reported in soybean, pea, and rice. This work will help to better understand the genetic architecture of seed yield in dry bean, and will assess the potential to combine MQTL, GWAS and comparative genomics to identify and refine the most stable MQTL-YC regions for their use in plant breeding.

Materials and Methods

Seed Yield and Yield Components QTL and QTN

A detailed literature search was carried out on common bean QTL and GWAS related to yield and yield components traits under sufficient water and drought conditions from 2000 to 2021. All QTL and GWAS except those lacking proper QTL related information, genetic map information, and those reported under other stress different than drought were used in the MQTL and QTN co-localization analyses. Based on these criteria, 394 QTL and 349 QTN for seed yield (YDSD) and six yield component traits including days to flower (DF), days to maturity (DMP), harvest index

(HI), pod harvest index (PHI), pods per plant (PDPL), seed per pod (PDPL), and seed weight (SW) were identified from 24 biparental, 1 multiparent and 10 diversity panels of common bean from 21 QTL and 11 GWAS, including Andean, Middle American, and wild germplasm (Table 2.1). The crop ontology for agricultural data for common bean was used as reference to unify the name of traits (https://cropontology.org/term/CO_335:ROOT).

Table 2.1 Summary of QTL and GWAS used in the MQTL analysis for yield and yield components in dry bean.

Reference	Germplasm	Gene pool	size	Env	QTL-QTN	Analysis
Blair et al 2006	Cerinza x G24404	AxW	157	N	21	QTL
Wright & Kelly 2011	Jaguar x 115M	M	96	N	3	QTL
Mkwaila et al 2011	Tacana x PI318695	MxW	30	N	1	QTL
Mkwaila et al 2011	Tacana x PI313850	MxA	30	N	6	QTL
Galeano et al 2012	DP	A	80	D-N	60	GWA
Asfaw et al 2012	BAT477 x DOR364	M	97	D-N	7	QTL
Blair et al 2012	BAT477 x DOR364	M	113	D-N	10	QTL
Checa & Blair 2012	G2333 x G19839	MxA	84	N	3	QTL
Blair & Izquierdo 2012	Cerinza x G10022	AxW	138	N	8	QTL
Mukeshimana et al 2014	SEA5 x CAL96	MxA	125	D-N	30	QTL
Cichy et al 2014	Black Magic x Shiny Crow	M	100	N	5	QTL
Trapp et al 2015	Buster x Roza	M	140	D-N	27	QTL
Kamfwa et al 2015	ADP	A	237	N	9	GWA
Villordo-Pineda et al 2015	Pinto Villa x Pinto Saltillo	M	282	D-N	34	SMA
Moghaddam et al 2016	MDP	M	280	N	11	GWA
Hoyos-villegas et al 2016	Merlot x SER48/55/94	M	76,36,48	N	9	QTL
Hoyos-villegas et al 2017	SNAD	M	96	D-N	2	GWA
Heilig et al 2017	Puebla 152 x Zorro	M	122	N	8	QTL
Diaz 2018	BAT 881 X G21212	M	95	D-N	4	QTL
da Silva et al 2018	Ruda x AND277	MxA	376	N	16	QTL
Resende et al 2018	DP	M-A	188	N	4	GWA
Sandhu et al 2018	BK004-001 x H68-4	M	85	N	17	QTL
Onziga et al 2019	Portillo x Red Hawk	A	97	D-N	3	QTL
Nabateregga et al 2019	BRB 191 x SEQ 1027	A	128	D-N	7	QTL
Berny Mier et al 2019	ICA Bunsí x SXB405	M	226	D-N	122	QTL
Geravandi et al 2020	Goli x AND1007	MxA	100	N	13	QTL
Wu et al 2020	SCAAS	M-A	683	N	83	GWA
Keller et al 2020	VEF	A	481	D-N	19	GWA
Diaz et al 2020	MAGIC	M	636	D	42-50	QTL-GWA
Mir et al 2021	DP	M-A	96	N	34	GWA
Nkhata et al 2021	DP	M-A	99	N	43	GWA
Diaz et al 2022	SCR16xSMC40	M	100	D-N	10	QTL
Diaz et al 2022	SMC33xSCR16	M	100	D-N	6	QTL
Diaz et al 2022	SMC44xSCR9	M	100	D-N	16	QTL

Germplasm: andean diversity panel (ADP), middle american diversity panel (MDP), subset of north american diversity panel (SNAD), subset Chinese academy of agriculture sciences (SCAAS), diversity panel (DP), Vivero Equipo Frijol (VEF), multiparent advanced generation intercross (MAGIC). Gene pool: andean (A), middle american (M), wild (W). Environment: drought (D), Non-drought (N).

The molecular markers with the highest test statistics associated with QTL were regarded as the estimated location of QTL for each reported association. When the position of the peak markers was not reported, the flanking marker position was used. If the position of the peak marker were not reported and flanking markers were in a different chromosome the QTL were not used. When the physical location of markers were not reported in the studies, we search for the amplicon

sequences on the Legume Information System (Dash et al., 2016) (**Supplemental Table 2.1**). When the SNP physical position were reported on *P. vulgaris* v1 genome, we extracted the 600 bp surrounding sequence of *P. vulgaris* v1 (**Supplemental Table 2.2**). The amplicons and surrounding sequences were used for the Basic Local Alignment Search Tool (BLAST) analysis against the *P. vulgaris* v2.1 genome for detecting the physical position. The number of QTL and QTN associated with yield and yield components were visualized graphically via R package ggplot2 3.3.5 (Wickham, 2009).

Conversion of physical to cM positions

The estimated location for all SNP markers in BARCBean6K_1 and BARCBean6K_2 chips, and QTL and QTN were converted to cM position in the Stampede x Red Hawk reference map based on physical position for the molecular markers on the *P. vulgaris* v2.1 genome (Schmutz et al., 2014; Song et al., 2015) using a special version of cM converter for common bean (<http://mapdisto.free.fr/cMconverter/>). For the projection of QTL, the confidence interval (CI) of 95% was estimated in the position where the molecular marker with the highest LOD value was reported for each QTL. The formulas were $CI = 530/(N \times R^2)$ for backcross (BC), and $CI = 163/(N \times R^2)$ for recombinant inbred line (RIL) populations (Guo et al., 2006a), where N is the population size and R^2 is the proportion of phenotypic variance of the QTL. If the CI were beyond the end of a chromosome, the CI was cut off at the end of that chromosome. The QTN were projected into the reference map to allow the comparison of GWAS and the MQTL analysis.

MQTL Analysis and QTN co-localization

The MQTL analysis for yield and yield components (MQTL-YC) was conducted in BioMercator v4.2 software (Sosnowski et al., 2012a), and the best model of MQTL-YC was chosen according to the prevalent value among Akaike Information Criterion (AIC), corrected Akaike Information criterion (AICc and AIC3), Bayesian Information Criterion (BIC) and Average Weight of Evidence (AWE) criteria. The CI of the MQTL-YC was defined as the most likely region but when QTL that belong to MQTL-YC were out of the CI, we used the extreme QTL peaks as boundaries for further analyses. QTL, QTN, and MQTL-YC were visualized graphically via Circos (Krzywinski et al., 2009)

Ortho-MQTL analysis

The *P. vulgaris* v2.1 (Schmutz et al., 2014) genome was compared to the *Glycine max* Wm82.a2.v1 (Schmutz et al., 2010), *Pisum sativum* v1a (Kreplak et al., 2019) and *Oryza sativa* v7

(Ouyang et al., 2007) genomes using the phyton version of MCScan (<https://github.com/tanghaibao/jcvi/wiki/MCscan>). To detect ortho-MQTLs between common bean and the other crops, we used the MQTL and MGWA analysis reported for soybean (Shook et al., 2021), pea (Klein et al., 2020) and rice (Khahani et al., 2021) to identify the physical position on the genomes of regions associated with seed yield and seed yield-components and we filtered out the syntenic blocks that were out of MQTL-YC identified on common bean. The EnsemblPlants database (Bolser et al., 2016) was used to identify the candidate and orthologous genes among species and the paralogous genes in common bean.

Results

Distribution of yield and yield-components QTL and QTN

Hundreds of molecular markers related to seed yield and yield components have been reported in common bean. To uncover the most consistent genomic regions associated with seed yield and seven yield components under sufficient water and drought conditions, a total of 394 QTL from 24 biparental and one multi-parental population were used. The populations included two Andean, and 15 Middle American intra-gene pool, five inter-gene pool, two Andean x wild, and a Middle American x wild populations. These populations were field grown under sufficient water and/or drought conditions from 1999 to 2017 and were reported in 21 studies (**Table 2.1, Supplemental Table 2.3**). From the 394 QTL, 223 were identified under sufficient water and 171 under drought conditions. The QTL reported for seed weight (SW), days to flowering (DF), yield (YDSD), and pod harvest index (PHI) were the most common QTL reported in the studies representing 34%, 20%, 16% and 13% of total QTL, respectively.

Furthermore, a total of 349 QTN were compiled for seed yield and seven yield components from 10 GWAS and one biparental population study that used a simplified method derived from GWAS, a Single-Marker Analysis (SMA). The SMA and GWAS included germplasm of Andean and Middle American gene pools. The populations were field grown under sufficient water and/or drought conditions from 2009 to 2019 and were reported in 11 studies. (**Table 2.1, Supplemental Table 2.4**). From the 349 QTN, 217 were identified under sufficient water, 79 under drought, and 53 under combined analysis of both conditions. The QTN reported for DF, DPM, SW, YDSD were the most common representing 31%, 17%, 17%, and 14% of total QTN, respectively.

The QTL, SMA and GWAS were conducted at 35 different locations distributed in 13 countries around the world (**Figure 2.1**). The distribution of QTL and QTN were unevenly distributed across

the eleven chromosomes of common bean (**Figure 2.2**). Chromosome Pv01 presented the highest number of QTL (69) and QTN (56) and chromosome Pv10 and Pv09 have the lowest number of QTL and QTN (15) (**Figure 2**).

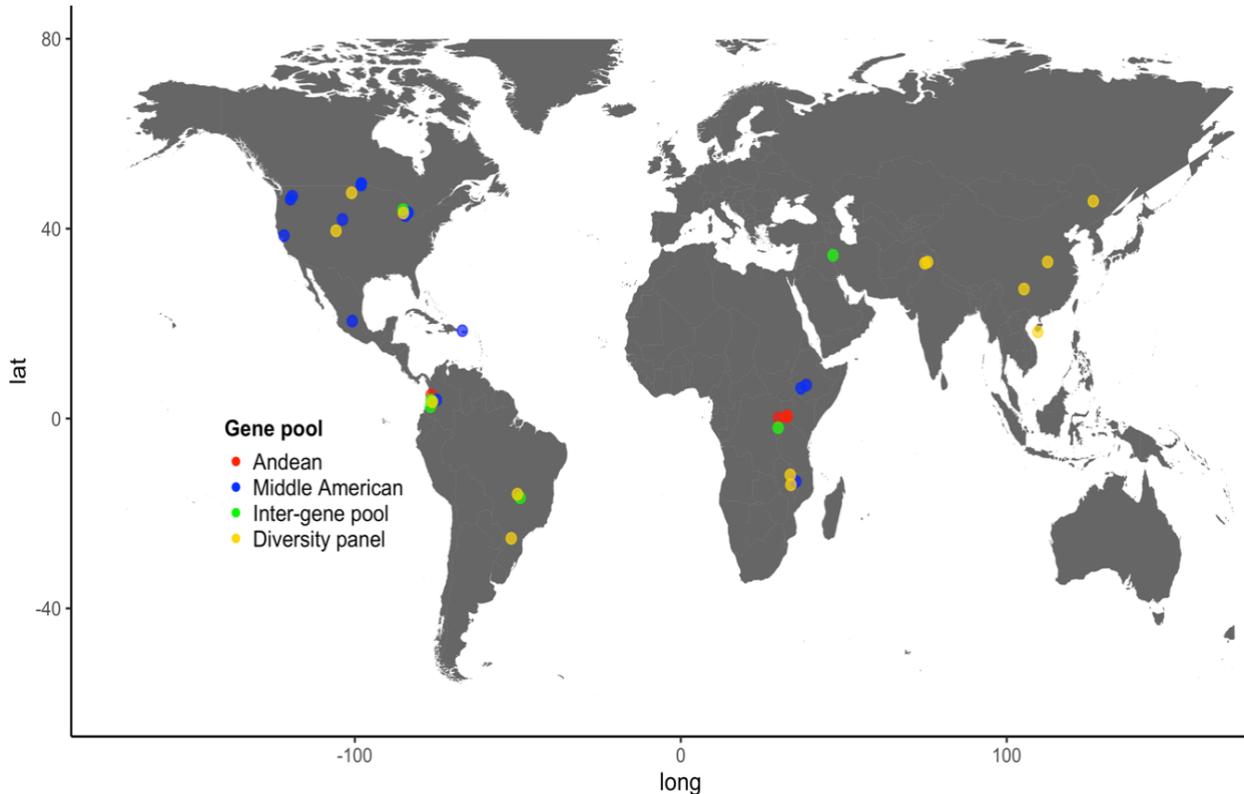


Figure 2.1 Geographic distribution of QTL and GWAS used for the MQTL and co-localization analyses.

Meta-QTL analysis

The physical position of 394 initial QTL were converted to cM using the recombination estimated in the Stampede x Red Hawk map. Stampede x Red Hawk map has a total length of 1042.2 cM and an average distance among SNP of 0.17 cM (Song et al., 2015). The MQTL analysis confined 373 (95%) of initial QTL into 58 MQTL for yield and yield components (MQTL-YC) supported by at least two QTL identified in different populations, environments, or traits (**Table 2.2, Supplemental Table 2.5**). The number of MQTL-YC per chromosome ranged from three on chromosome Pv10 to eight on chromosome Pv01 (**Figure 2.3**). The Meta-QTL analysis allowed a 3.2 reduction on average of the CI (3.8 cM) in comparison to the average CI of initial QTL (12 cM). The MQTL-YC with the largest physical sizes were MQTL-YC1.2 (14.7 Mb),

MQTL-YC3.3 (21.3 Mb), MQTL-YC4.2 (11.6 Mb), MQTL-YC6.1 (8.52 Mb), MQTL-YC7.5 (28.51 Mb) and MQTL-YC8.3 (42.2 Mb), all of them within the pericentromeric regions

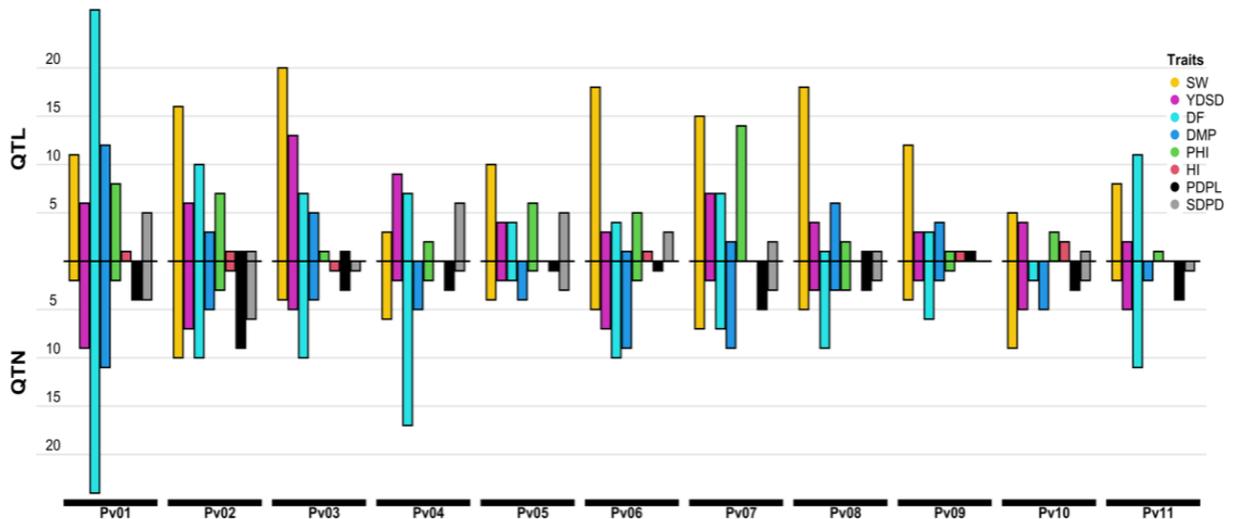


Figure 2.2 The number and distribution of QTL and QTN associated with yield and yield components on common bean.

Further support for 40 MQTL-YC was obtained by the co-localization of 210 QTN reported for seed yield and seed yield components in common bean. Since most of the QTN do not fall into coding regions, and the linkage disequilibrium can affect the co-localization, QTN were considered co-localized if they were within the CI of MQTL-YC or within the boundaries of QTL peaks that support each MQTL-YC (**Table 2.2, Supplemental Table 2.6**). The co-located QTN led to the identification of 42 candidate genes that have been related with flowering, circadian clock, root elongation, plant shoot branching, plant growth, leaf development, photoperiod, seed weight, seed development, seed yield and seed yield components in several species as common bean, arabidopsis, rice, soybean, ryegrass, and maize (**Supplemental Table 2.7**). The well-known *fin* (Phvul.001G189200) and *Ppd* (Phvul.001G221100) genes are in MQTL-YC1.6 and MQTL-YC1.7, respectively, and several QTN were reported around them, which show the potential of QTN to narrow down genomic regions to identify candidate genes.

MQTL-YC across yield components

Multiple yield component QTL/QTN were identified within individual MQTL-YC (**Figure 2.4**). In total, 38 out of 58 MQTL-YC included QTL/QTN for YDSD. The co-localization of QTL/QTN for seed yield components with YDSD ranged from 68% (PHI) to 100% (HI). The QTL/QTN for DF, DPM, PHI, SW and YDSD most frequently co-localized within MQTL-YC,

however this could be a result of the higher number of QTL/QTN reported for these traits as compared to HI, PDPL and SDPD (**Figure 2.4**). Pods per plant (PDPL) has been reported to be critical for YDSD in dry bean (Kelly, 2018). QTL/QTN for PDPL had 93% co-localization with SW and 53% with SDPD.

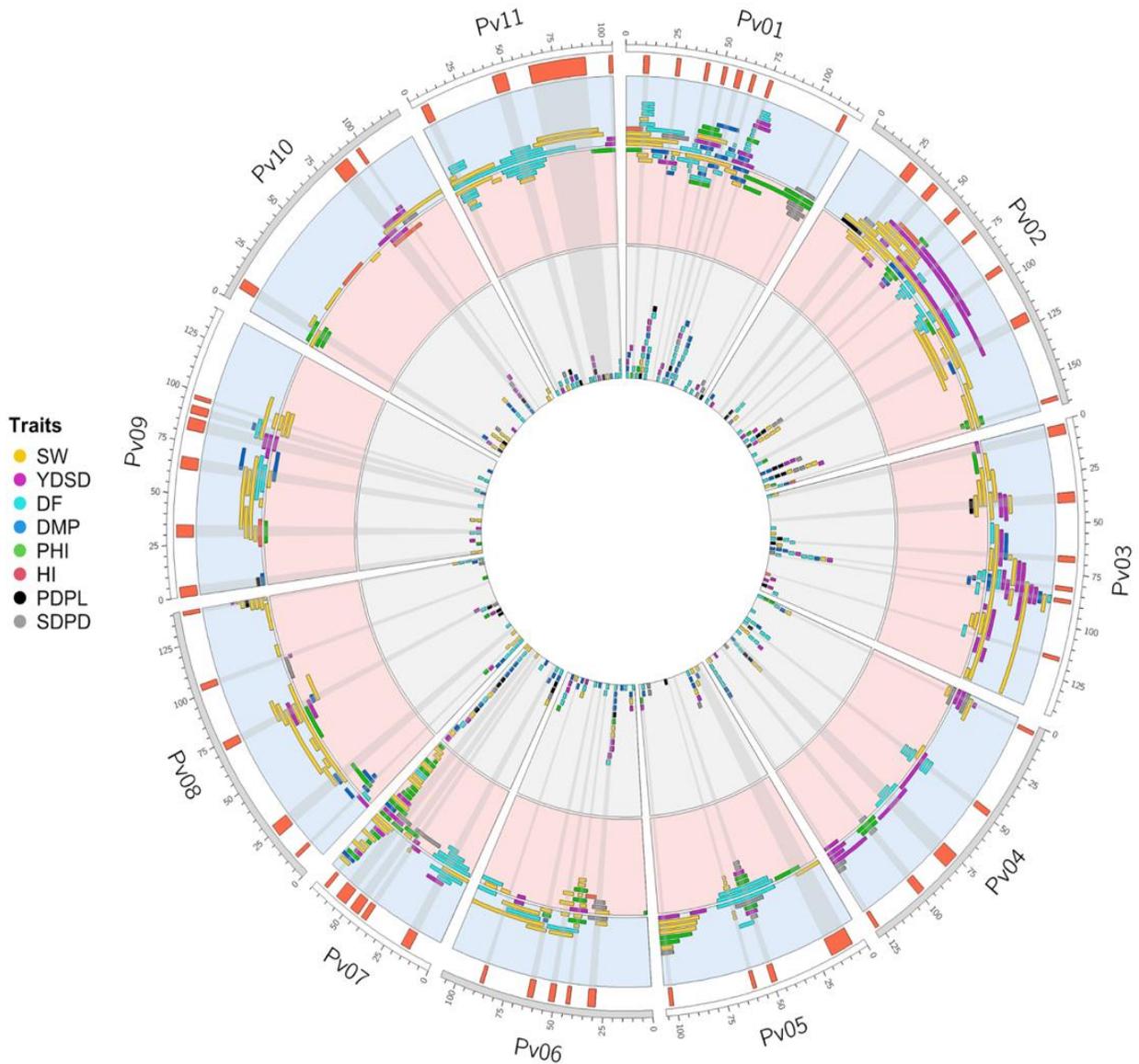


Figure 2.3 Circus plot showing distribution of MQTL-YC, QTL and QTN associated with yield and yield components on common bean. The outermost circle indicates the length on cM on the reference genetic map. The second circle indicates the MQTLs with 95% confidence.

Table 2.2 Description of detected MQTL-YC.

MQTL	Pos (cM)	CI (cM)	Pos (Mb)	QTL-QTN	N of populations	QTL-QTN	Traits	Env	GP	Lat
MQTL-YC1.1	10.4	2.4	2.4	12-7	7-3		DF, DPM, HI, PHI, SW, YDSD	D6, N13	B	B
MQTL-YC1.2	27.4	0.6	11.4	11-16	4-6		DF, DPM, PDPL, SDPD, SW, YDSD	D15, N10, C2	B	B
MQTL-YC1.3	41.9	1.3	29.8	5-0	1-0		DF	D3, N2	M	Temp
MQTL-YC1.4	51.0	2.0	37.1	11-3	5-2		DF, DPM, PHI, SW, YDSD	D7, N7	B	B
MQTL-YC1.5	58.5	3.7	42.8	4-1	1-1		DF, DPM	D2, N3	M	Trop
MQTL-YC1.6	65.8	1.2	45.0	4-6	3-2		DF, DPM, SW, YDSD	D1, N9	M	B
MQTL-YC1.7	74.5	1.2	47.2	16-6	4-2		DF, DPM, PHI, SW, YDSD	D4, N18	B	B
MQTL-YC1.8	115.7	1.9	51.2	5-3	2-2		PHI, SDPD	D4, N4	M	B
MQTL-YC2.1	27.2	5.0	1.8	5-0	3-0		PDPL, SDPD, SW	D2, N3	M	B
MQTL-YC2.2	41.8	3.5	2.9	8-2	5-2		DF, HI, SW, YDSD	D2, N7, C1	B	B
MQTL-YC2.3	58.6	2.4	8.3	7-0	2-0		DPM, PHI, SW, YDSD	D4, N3	M	Temp
MQTL-YC2.4	71.9	2.4	71.8	5-0	1-0		DF	D3, N2	M	Temp
MQTL-YC2.5	93.5	3.7	30.7	6-12	3-4		DF, PDPL, PHI, SDPD, SW, YDSD	D5, N13	B	B
MQTL-YC2.6	120.5	5.3	41.0	5-2	2-2		DF, DPM, PHI, SW	D5, N2	M	B
MQTL-YC2.7	161.4	0.4	47.7	5-2	2-1		PHI, SW	D5, N2	M	B
MQTL-YC3.1	4.8	5.4	1.0	3-1	3-1		DF, PHI, SW, YDSD	D2, N2	M	B
MQTL-YC3.2	37.4	4.4	3.6	9-0	2-0		PDPL, SW, YDSD	D4, N5	B	B
MQTL-YC3.3	67.5	3.2	31.4	5-5	4-5		DF, DPM, PDPL, SW, YDSD	D1, N9	B	B
MQTL-YC3.4	82.3	2.0	39.6	9-11	5-4		DF, DPM, SDPD, SW, YDSD	D8, N12	B	B
MQTL-YC3.5	88.4	2.2	41.7	10-1	6-1		DF, DPM, SW, YDSD	D1, N9, C1	B	B
MQTL-YC3.6	115.7	0.8	46.4	8-3	5-2		DF, HI, SW, YDSD	D3, N8	B	B
MQTL-YC4.1	4.2	2.6	0.5	7-6	6-5		DF, DPM, PHI, SDPD, SW, YDSD	D4, N8, C1	B	B
MQTL-YC4.2	48.5	3.0	12.6	7-5	3-3		DF, PDPL, SW, YDSD	D4, N8	B	B
MQTL-YC4.3	77.9	7.7	41.2	3-7	2-3		DF, DPM, PDPL, SW, YDSD	D4, N6	B	B
MQTL-YC4.4	98.9	3.4	43.6	4-0	1-0		PHI, SDPD	D1, N3	M	Temp
MQTL-YC4.5	127.8	1.0	45.4	6-6	4-2		DF, DPM, SDPD, YDSD	D1, N10, C1	B	B
MQTL-YC5.1	11.6	12.0	1.0	2-4	2-3		DPM, PHI, SW	D1, N4, C1	B	B
MQTL-YC5.2	49.3	3.3	4.7	13-0	3-0		DF, PHI, SDPD, YDSD	D7, N6	B	B
MQTL-YC5.3	59.9	2.8	12.4	5-0	1-0		SW	D3, N2	M	Temp
MQTL-YC5.4	102.6	1.5	39.2	8-2	4-1		DF, DPM, PHI, SDPD, SW, YDSD	N8, C2	B	B
MQTL-YC6.1	30.8	3.4	12.2	6-11	3-3		DF, DPM, HI, PDPL, SDPD, SW, YDSD	D10, N7	B	B
MQTL-YC6.2	43.0	1.7	18.7	10-1	2-1		DF, PHI, SW, YDSD	D7, N4	M	B
MQTL-YC6.3	51.6	3.4	19.6	5-0	3-0		DF, SW	D2, N3	B	Trop
MQTL-YC6.4	62.6	2.3	21.0	4-0	4-0		DPM, PHI, SW	N4	B	Trop
MQTL-YC6.5	87.3	0.5	26.8	9-11	4-5		DF, DPM, PHI, SW, YDSD	D7, N11, C2	B	B
MQTL-YC7.1	15.4	4.8	1.9	8-3	4-2		DF, DPM, SW, YDSD	D5, N6	M	B
MQTL-YC7.2	40.3	2.6	5.5	4-0	4-0		SDPD, SW, YDSD	D2, N2	B	B
MQTL-YC7.3	46.3	5.1	7.2	2-1	2-1		PHI, SW	D2, C1	B	Trop
MQTL-YC7.4	55.0	6.8	8.9	2-0	1-0		PHI, SDPD	D2	M	Trop
MQTL-YC7.5	63.9	0.3	27.0	31-20	5-8		DF, DPM, PDPL, PHI, SDPD, SW, YDSD	D23, N22, C6	B	B
MQTL-YC8.1	9.4	1.9	1.0	7-5	4-4		DF, DPM, PHI, SW, YDSD	D6, N6	M	B
MQTL-YC8.2	25.2	5.7	3.1	2-0	2-0		DPM, SW	N2	M	Temp
MQTL-YC8.3	72.7	3.7	44.5	9-7	6-4		DF, DPM, PDPL, PHI, SDPD, SW, YDSD	D6, N10	B	B
MQTL-YC8.4	103.6	2.3	54.0	3-0	3-0		SDPD, SW, YDSD	D2, N1	M	Trop
MQTL-YC8.5	141.3	0.0	59.5	9-10	6-6		DF, DPM, PDPL, SDPD, SW, YDSD	D2, N16, C1	B	B
MQTL-YC9.1	2.6	4.3	1.9	2-2	2-1		DPM, PDPL, SW	N4	A	B
MQTL-YC9.2	31.8	6.2	14.3	4-3	3-2		DF, HI, PHI, SW, YDSD	D1, N6	B	B
MQTL-YC9.3	64.9	5.6	23.7	7-1	4-1		DF, DPM, SW	D2, N6	B	B
MQTL-YC9.4	84.2	5.8	26.6	3-1	2-1		DF, YDSD	D1, N2, C1	B	B
MQTL-YC9.5	90.8	3.5	29.5	2-0	1-0		DF, DPM	N2	A	Trop
MQTL-YC9.6	97.4	1.8	31.1	5-0	1-0		SW	D3, N2	M	Temp
MQTL-YC10.1	9.5	4.7	2.7	5-4	3-3		DPM, PDPL, PHI, SW, YDSD	D3, N5, C1	M	B
MQTL-YC10.2	85.8	8.4	40.9	3-7	3-2		DPM, PDPL, SDPD, YDSD	D3, N7	B	B
MQTL-YC10.3	96.9	1.0	41.7	3-0	3-0		HI, SDPD, YDSD	D1, N2	B	B
MQTL-YC11.1	5.9	3.3	0.6	9-2	3-2		DF, SW, YDSD	D5, N6	B	B
MQTL-YC11.2	46.5	6.3	5.7	6-0	2-0		DF, SW	D4, N2	M	B
MQTL-YC11.3	76.5	29.2	38.7	2-8	1-6		DF, DPM, PDPL, SW, YDSD	D3, N6, C1	B	B
MQTL-YC11.4	104.7	1.2	49.7	3-2	2-2		DF, PHI, YDSD	D2, N3	M	B

Environment (Env): Drought (D), non-drought, and combined (C) conditions. Germplasm (GP): MQTL-YC with effects in Andean (A), Middle American (M), and both (B) gene pools. Latitude (Lat): MQTL-YC within effects in temperate (Tem), tropical (Trop), and both latitudes.

Interestingly, the co-localization of QTL/QTN for PDPL with QTL/QTN for partitioning traits (HI (7%), PHI (27%)) was lower compared to SDPD (HI (33%), PHI (47%)) and SW (HI (83%), PHI (80%)). QTL/QTN for phenology traits (DF, DPM) showed a high co-localization frequency with the other traits, ranging for DPM from 53% (SDPD) to 87% (SW).

The co-localization of different traits could be explained by pleiotropy or linkage disequilibrium among causative loci. Although it is challenging to distinguish between them (Chebib and Guillaume, 2021), some MQTL-YC seems to have pleiotropy caused by major QTL, for example MQTL-YC1.6 comprising the *fin* locus that control growth habit (**Figure 2.5a**), while others such as MQTL-YC3.5 seems to be controlled by tightly linked loci (**Figure 2.5b**). MQTL-YC1.6 and MQTL-YC3.5 are both associated with DF, DPM, YDSD and SW. However, in MQTL-YC1.6 all the QTL/QTN related to DF, DPM and YDSD are surrounding the *fin* loci, while in MQTL-YC3.5 the QTL are clustered by trait.

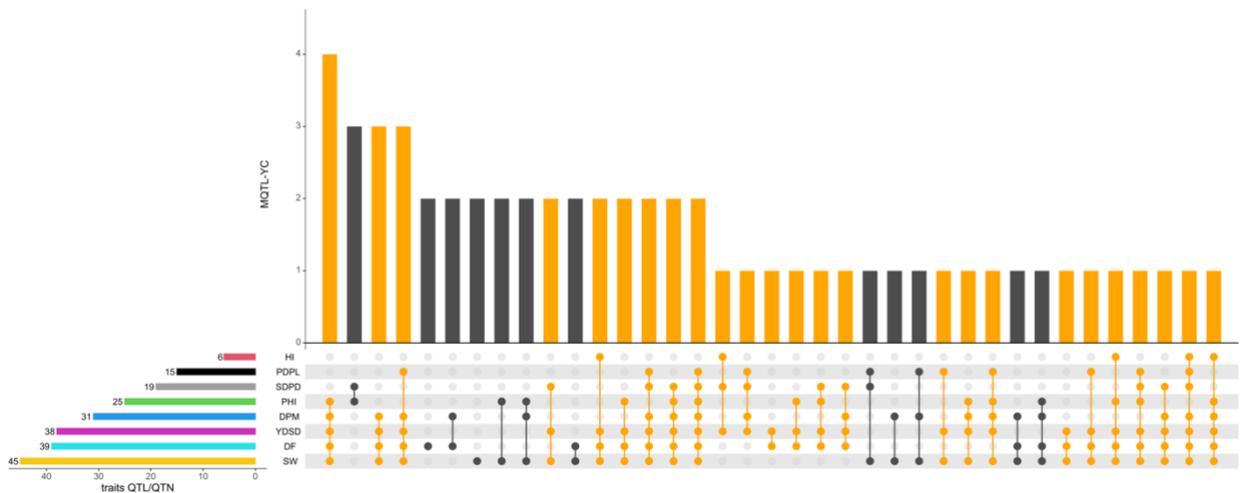


Figure 2.4 Upset plot showing the interaction among traits within the 58 MQTL-YC. The number of MQTL-YC in which each trait was reported are displayed as horizontal bars in the lower-left corner. The number of intersections are shown as vertical bars, with those including YDSD QTL/QTN highlighted in orange. The QTL/QTN traits involved in each intersection are identified with connected circles, with those including YDSD QTL/QTN highlighted in orange.

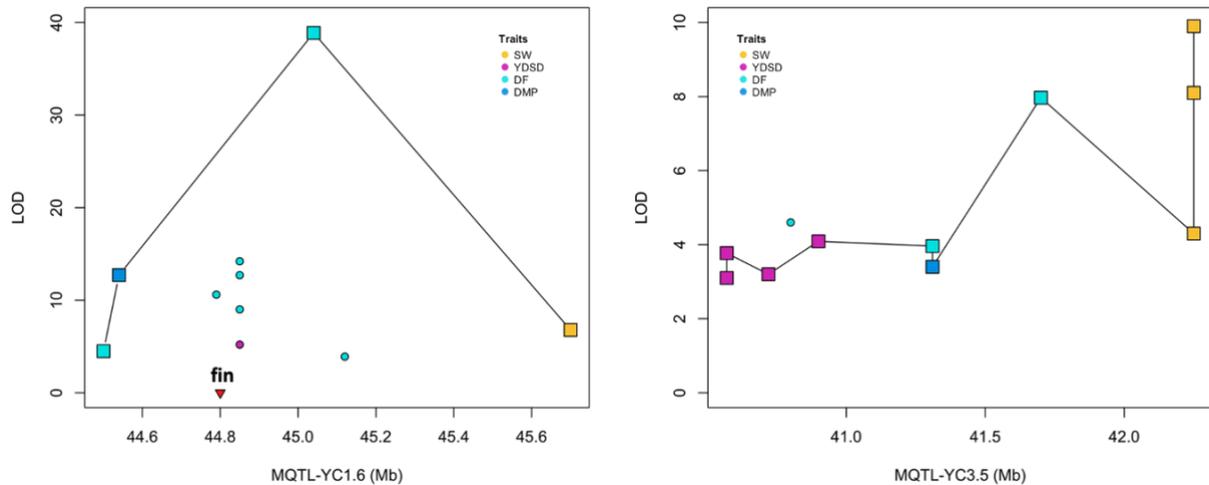


Figure 2.5 Co-localization of QTL and QTN in a) MQTL-YC1.6 and b) MQTL-YC3.5. The squares and circles represent QTL and QTN, respectively. The red triangle in MQTL-YC1.6 is the location of the fin locus.

MQTL-YC across gene pool and latitude

From the 373 QTL contained in the MQTL-YC, the parental sources originated from wild, Andean, and Middle American gene pools for 9, 68, and 296 QTL, respectively. Additionally, from the 210 QTN that co-located with MQTL-YC, 53 and 57 were reported in Andean and Middle American populations, respectively. The remainder 100 QTN were reported in diversity panels including both gene pools. In total, 327 associations were identified in tropical (<30 degrees north latitude), and 253 in temperate (>30 degrees north latitude) regions. Out of the total number of MQTL-YC, two were specific for Andean, 22 for Middle American, and 34 have sources from both gene pools. In total, 31 MQYL-YC were supported by QTL/QTN identified in both gene pools and latitudes, which suggests that these regions could have an effect on a wide genetic background, independently of the latitude (**Table 2.2**).

MQTL-YC across gene pool and latitude

From the 373 QTL contained in the MQTL-YC, the parental sources originated from wild, Andean, and Middle American gene pools for 9, 68, and 296 QTL, respectively. Additionally, from the 210 QTN that co-located with MQTL-YC, 53 and 57 were reported in Andean and Middle American populations, respectively. The remainder 100 QTN were reported in diversity panels including both gene pools. In total, 327 associations were identified in tropical (<30 degrees north latitude), and 253 in temperate (>30 degrees north latitude) regions. Out of the total number of MQTL-YC, two were specific for Andean, 22 for Middle American, and 34 have sources from

both gene pools. In total, 31 MQYL-YC were supported by QTL/QTN identified in both gene pools and latitudes, which suggests that these regions could have an effect on a wide genetic background, independently of the latitude (**Table 2.2**).

MQTL-YC related with drought

In total, 51 out of 58 MQTL-YC were supported by QTL/QTN identified under both non-drought and drought conditions, which supports the hypothesis that breeding for both conditions simultaneously is possible. However, five out of those 51 (MQTL-YC1.4, MQTL-YC6.1, MQTL-YC6.5, MQTL-YC7.5 and MQTL-YC8.3) appear to be relevant for drought due to the high number (>6) of QTL/QTN in this environment and their association with partitioning. All these MQTL-YC were supported by QTL/QTN identified in both gene pools and latitudes (**Supplemental Table 2.8**).

Ortho-MQTL analysis

Genomic synteny was evaluated among the MQTL-YC identified in dry bean and MQTL studies for pea (Klein et al., 2020) and rice (Khahani et al., 2021) and a MGWAS study for soybean (Shook et al., 2021). In total, 43 MQTL-YC identified in this study showed synteny blocks (ortho-MQTL-YC) with genomic regions associated with seed yield components in pea, soybean, and/or rice (**Supplemental Table 2.9 and 2.10**). The most consistent ortho-MQTL-YC were identified across legumes (**Table 2.3, Figure 2.6**). Among the ortho-MQTL-YC, the MQTL-YC3.6 was the only one that showed synteny with genomic regions associated with yield components in rice, soybean, and pea. Interestingly, the gene Phvul.003G252400 (49.2 Mb) located in MQTL-YC3.6 has orthologous genes related with the ortho-MQTL-YC in rice (LOC_Os02g45054) and soybean (Glyma.16G141100). The LOs02g45054 gene has been related with the transition from vegetative to reproductive growth in rice (Deng et al., 2017). In total, 38 out of the 43 ortho-MQTL-YC showed signatures of selection in common bean (Schmutz et al., 2014; Wu et al., 2020a), which could indicate a convergent evolution of domestication traits in legumes (Wang et al., 2018; Rau et al., 2019).

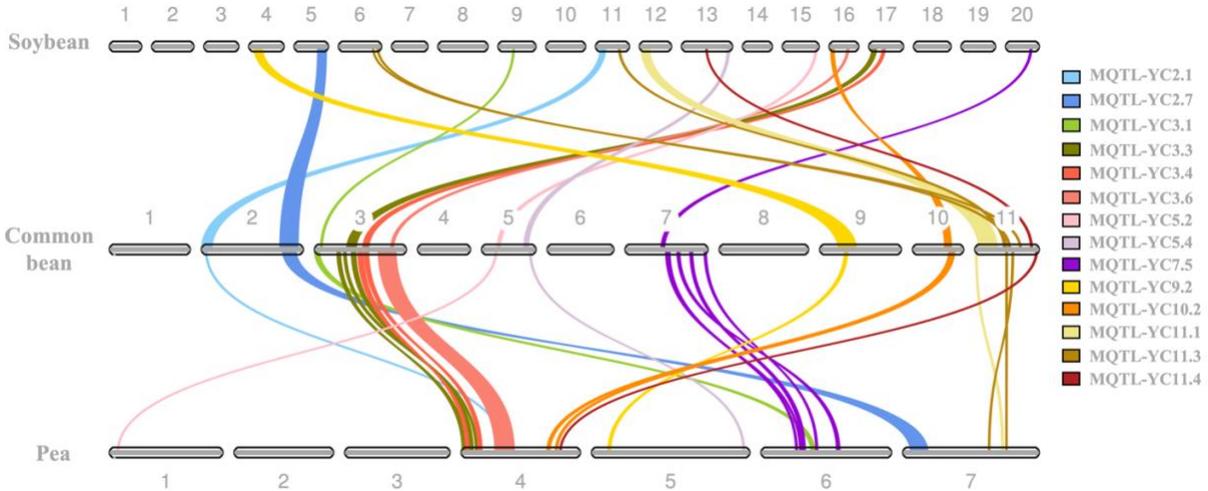


Figure 2.6 Comparative maps of ortho-MQTLs among common bean, soybean and pea.

Table 2.3 Ortho MQTL-YC detected between common bean, soybean and pea.

MQTL	Start (Mb)	End (Mb)	QTL-QTN	GP
MQTL-YC2.1	1.4	1.9	5-0	M
MQTL-YC2.7	46.8	47.7	5-2	M
MQTL-YC3.1	0.2	2.0	3-1	M
MQTL-YC3.3	11.7	33.0	5-5	B
MQTL-YC3.4	35.4	40.0	9-11	B
MQTL-YC3.6	44.3	49.2	8-3	B
MQTL-YC5.2	4.6	7.2	13-0	B
MQTL-YC5.4	38.7	39.5	8-2	B
MQTL-YC7.5	11.1	39.6	31-20	B
MQTL-YC9.2	14.0	15.4	4-3	B
MQTL-YC10.2	40.5	41.3	3-7	B
MQTL-YC11.1	0.4	1.3	9-2	B
MQTL-YC11.3	10.1	45.4	2-8	B
MQTL-YC11.4	49.4	51.8	3-2	M

Germplasm (GP): MQTL-YC with effects in Middle American (M), and both (B) gene pools. *Maturity date (MD).

Discussion

Seed yield and seed yield components QTL and QTN on common bean

QTL and GWAS studies provide rich datasets to understand the genetic architecture of seed yield in common bean and to use that information to improve breeding methods and targets. The value of these studies is greatly increased when they are taken collectively as with the MQTL approach. The integration of QTL and GWAS is an effective approach to take advantage of the power of detection of QTL and high resolution of GWAS, narrowing down the number of

candidate genes (Brown et al., 2021; Bilgrami et al., 2022). Seed yield is a complex trait that is controlled for many QTL with small effects, and it is strongly affected by environmental conditions. This complex genetic architecture reduces the power to uncover the genomic regions that are associated with seed yield. To address this limitation, we identified seed yield components that have been widely used in QTL and GWAS that have a less complex genetic architecture controlled by fewer genes. Disaggregating complex traits into more simple ones can increase the power to detect the causative loci (Benjamin et al., 2012).

The physical positions of all QTL and QTN evaluated in this study were physically positioned in relation to the *P. vulgaris* reference genome v2.1. This allowed the comparison of genetic maps from *i*) different mapping populations that depend on the number of markers and recombination events to estimate the genetic positions of markers, and *ii*) studies that reported the QTL/QTN positions on reference genome v1, since the physical position of markers change in the most updated reference genome v2.1. The identification of physical positions in v2.1. for all QTL (394) and QTN (349) included in this study allowed for the comparison from independent experiments and environments and made an ideal dataset for MQTL analyses due to the high reliability of the physical markers position on the latest genome reference v2.1.

MQTL analysis

The current MQTL analysis included QTL and QTN identified in the last 20 years in dry bean. However, not all the QTL/QTN reported for seed yield components were considered in the analysis because lack of information to identify their physical position. Additionally, instead of using a consensus map that depends on common markers among individual mapping populations to project the QTL/QTN, we used the recombination rate estimated in the highly saturated Stampede x Red Hawk reference map genotyped with BARCBean6K_1 and BARCBean6K_2 BeadChips that are commonly used for QTL and GWAS in dry bean (Song et al., 2015; Moghaddam et al., 2016; Heilig et al., 2017; Geravandi et al., 2020). The use of recombination rate in the Stampede x Red Hawk reference map *i*) overcome the need for common markers among individual maps, and *ii*) improve the quality of the MQTL because the population size (267) is up to 8-fold compared to the populations included in this study, which translate in a better estimation of the recombination.

This study narrows down the CI of detected MQTL-YC compared to the initial QTL. However, although the gaussian mixture model implemented in BiomeRCator have been prove to be a good

clustering approach to determine the real number of distinct QTL (Veyrieras et al., 2007; Sosnowski et al., 2012b), factors as population size, number of markers and QTL effect influence the CI of QTL, which affects the CI of the MQTL (Visscher and Goddard, 2004; Guo et al., 2006b). Moreover, other biological factors such as linkage disequilibrium and genomic regulatory elements that can influence gene expression of causative genes could be underestimated by the CI of the MQTL. For the above, although the CI is the most likely region estimated by the gaussian mixture model, we were conservative and when QTL were out of the CI, we used the extreme QTL to define the MQTL-YC regions. The MQTL-YC with the largest physical sizes were in the pericentromeric regions. The pericentromeres contained 26.5% of the genes identified in common bean, and the average recombination rate is lower (5.1 Mb/cM) compared to euchromatic arms (0.24 Mb/cM) (Schmutz et al., 2014), which explained the large physical size of these MQTL-YC, especially for MQTL-YC7.5 and MQTL-YC8.3 regions with a recombination rate of 9.2 and 6.9 Mb/cM, respectively (Schmutz et al., 2014).

Additional support for MQTL-YC were assessed with the co-localization of QTN associated with seed yield and seed yield components. In this study, 60% (210) of QTN included co-localized with 40 MQTL-YC, which suggest that 40% of QTN could be *i*) false-positive associations related with population structure or *ii*) QTN related with rare alleles that are not common in breeding programs. Interestingly, candidate genes associated with selection signatures have been reported on 49 out of 58 MQTL-YC genomic regions in previous studies (Schmutz et al., 2014; Wu et al., 2020b) (**Supplemental Table 2.11**).

The validation of candidate genes in dry bean is challenging because the lack of an efficient transformation methodology (Hnatuszko-Konka et al., 2014). However, promising candidate genes were identified by integrating MQTL-YC and reported QTN. The well-known genes *fin* and *Ppd* related with determinacy and photoperiod sensitivity were identified precisely within MQTL-YC with several QTN around them (<10 kb). We found that focusing on QTN that have co-localization with MQTL-YC is a powerful approach to narrow down the genomic regions leading to the identification of 42 candidate genes. The receptor associated kinase (RAK) Phvul.006G020700 was closely located to QTN (69 kb) reported for YDSD, SW and DPM in MQTL-YC6.1. In a study, transgenic plants silenced for RAK presented a reduction of seed number in rice (Zhou et al., 2016). Interestingly, three QTL for SDPD were detected in MQTL-YC6.1, which suggest that the role of Phvul.006G020700 could be similar in dry bean. MADS-

box protein have been related with seed development in Arabidopsis (Gramzow and Theissen, 2010), and three MADS-box protein belonging to MQTL-YC2.7 are closely located to QTN for SW and PHI. In addition, MQTL-YC2.7 comprised QTL for SW and PHI, which provide evidence that this region is related with seed development. Although the 42 candidates genes detected within MQTL-YC can potentially have the same function of orthologous genes, they should be validated, and in absence of an efficient transformation methodology, functional genomics is an alternative to provide this validation in future studies (Ibrahim et al., 2020).

MQTL-YC across yield components

Finding that 51 out of 58 MQTL-YC were supported by QTL/QTN identified under both drought and non-drought conditions supports the hypothesis that yield potential is not exclusive between non-stressful and stressful environments (Beebe et al., 2008; Diaz et al., 2018), and genes related to traits such as photosynthate translocation, phenology, plant architecture, seed and development could help to increase seed yield in both conditions. The MQTL-YC could be classified into two groups depending on whether they are related to phenology. 46 MQTL-YC are related to phenology (DF and/or DPM), and 12 have no QTL/QTN identified for DF or DPM. Overall, increasing days to maturity allows plants to produce more biomass and have more time for seed filling increasing partitioning, seed size and seed number leading to high seed yield under sufficient water conditions. However, under drought conditions, the tendency of positive correlation between seed yield with DF and DPM changes to negative due to abortion of flowers, pods and seeds caused for water stress, especially under terminal drought in drought susceptible lines. The antagonistic effect between DF and DPM with YDSD under drought conditions was observed with the susceptible (MIB778) and drought tolerant parents (SEN56 and ALB213) in a MAGIC population, where loci from MIB778 consistently increase DF and DPM, while loci of SEN56 and ALB213 increase SW and YDSD (Diaz et al., 2020b). Remarkably, some drought tolerant lines present increased yield in non-stressed conditions as well as slightly earlier maturity, suggesting genetic effects of crop efficiency that do not depend on longer vegetative period (Beebe et al., 2008). The previous finding is supported by five MQTL-YC (1.2, 2.3, 3.3, 6.5, 7.5) where QTL for YDSD and DF/DPM were identified in the same population and loci from drought-tolerant lines increased YDSD while reducing phenology traits (DF and/or DPM) (Wright and Kelly, 2011; Trapp et al., 2015; Berny Mier Y Teran et al., 2019; Diaz et al., 2020b). Additionally, phenology is affected by differences in photoperiod represented by long-day and short-day in

temperate and tropical regions, respectively. Photoperiod sensitivity is under oligogenic control, and has effects in the reproductive development (Weller et al., 2019). In a cross between an adapted and a photoperiod sensitive landrace, the long-day treatment reduced the internode length, increasing the number of pods per plant compared to short-day on dry bean (González et al., 2021). The differences between photoperiod sensitivity and reproductive development in high latitudes are related with local adaptation, and the higher number of pods reported by Gonzalez et al. (2021) could be beneficial for seed yield in temperate regions.

The most consistent MQTL-YC could be determined by the number of supported QTL/QTN and the number of different genetic backgrounds where the associations were reported. We used the following criteria to identify the most robust MQTL-YC in common bean: *i*) MQTL-YC with QTL/QTN in at least 5 populations, and *ii*) MQTL-YC with at least one QTL/QTN for YDSD. Based on these criteria, we identified 28 MQTL-YC that could be used in common bean breeding to increase seed yield (**Supplemental Table 2.8**). Interestingly, 27 of these MQTL-YC were supported by QTL/QTN reported with effect in both low and high latitudes and both water supplies, and from them, 23 were supported by QTL/QTN with both gene-pools as sources while the reminder four were supported by QTL/QTN where the sources were Middle American genotypes. The MQTL-YC5.4 was the only robust MQTL-YC that was supported for QTL/QTN under sufficient water conditions alone, which reinforced the hypothesis that genetic gain for seed yield is achievable for both drought and non-drought simultaneously.

These robust MQTL-YC could be used to deploy marker-assisted selection strategies. It is important to consider that all these MQTL-YC included phenological traits. However, QTL with high additivity for DF and DPM (>2 days) were in general reported in crosses *i*) between parents of different growth habit (Trapp et al., 2015), *ii*) including photoperiod sensitive material field grown in temperate regions (Heilig et al., 2017), and *iii*) including wild accessions (Blair et al., 2006), which suggest that increasing seed yield using these MQTL-YC in adapted germplasm with similar growth habit should not increase significantly DPM. Besides, breeders in temperate regions should look closely at the MQTL-YC1.7 comprising the *Ppd* locus controlling photoperiod sensitivity, which in a dominant state could increase DF and even produce non-flowering plants under long-day environments such in temperate regions (Weller et al., 2019).

Among the detected MQTL-YC in this study, 12 MQTL-YC are unrelated with phenology, and although they do not fulfill our criteria to be labeled as robust MQTL-YC, these MQTL-YC

have the potential to improve seed yield without increasing DF and DPM. As the number of different populations supporting each MQTL-YC reflects their potential use in a wide genetic background, we believe that the MQTL-YC unrelated with phenology with QTL/QTN identified in at least three populations could be considered for breeding decisions.

MQTL-YC related with drought

Drought is one of the most important abiotic stress in common bean around the world, and it is expected to affect more growing areas under future climate change scenarios (Beebe et al., 2008; Hummel et al., 2018). Under drought conditions, susceptible dry beans cultivars show slow vegetative growth and seed filling while increasing flower and pod abortion, and tolerant genotypes maintain high vegetative growth and seed filling, which suggests that high partitioning efficiency into seeds is a good indicator of drought tolerance (Hageman and Volkenburgh, 2021). In the populations used in this analysis, drought tends to reduce the number of pods and seeds per plant (Diaz et al., 2022), while increasing SW (Diaz et al., 2018), which could be explained by less competition for nutrients due to lower number of seeds during seed filling period.

Although the reported correlation between SW with PHI is not strong (Diaz et al., 2022), and sometimes negative (Kamfwa et al., 2015), we identified that 80% (25) of the MQTL-YC that involved QTL for PHI overlap with QTL reported for SW. Furthermore, a strong correlation of PHI and seed number has been reported (Berny Mier Y Teran et al., 2019; Diaz et al., 2022), indicating that sink strength is higher for the seed number than for SW. However, the number of seeds per pod is more sensitive to abiotic stresses such as drought and heat, reducing the number of seeds per pod (Berny Mier Y Teran et al., 2019; Diaz et al., 2022), which could lead to partitioning in a reduced number of seeds increasing their size (Mukeshimana et al., 2014; Diaz et al., 2020b). Although all the robust MQTL-YC but one identified in this study have QTL/QTN identified under drought conditions, five MQTL-YC appear to be more relevant for drought. These include, MQTL-YC1.4, MQTL-YC6.1, MQTL-YC6.5, MQTL-YC7.5 and MQTL-YC8.3 which originate from 6 to 23 QTL/QTN identified under drought. These MQTL-YC are related with phenology (DF, DPM) and partitioning (HI, PHI).

MQTL-YC related with domestication

Complex traits such as seed yield are controlled for many QTL with small effects across the genome. However, complex traits could have major QTL, but due to domestication, most of them are fixed (Doebley, 2006; Bernardo, 2008). Several major QTL for DF, DPM, PDPL and SW

belonging to MQTL-YC regions were reported in crosses that included landraces and wild germplasm (Blair et al., 2006; Blair and Izquierdo, 2012; Checa and Blair, 2012; Geravandi et al., 2020). The highest additivities for each trait were 11.4 grams for SW, four days for DF, three days for DPM, and five pods for PDPL. Taking into account that the definition of additivity is half of the difference between the two homozygous in biparental populations (Li et al., 2011), the effect of these regions is large. However, QTL for the same traits reported in crosses between adapted cultivars show that the effect is 10-fold lower, which reflects the effect of domestication in common bean.

Furthermore, we identified that 58% (34) of the MQTL-YC are supported by QTL identified in both gene pools. This could indicate that selection pressure for yield led to focusing on similar genomic regions in both gene pools. Schmutz et al. (2014) reported that less than 10% of the genes related with domestication were similar between both gene pools. However, this proportion could be underestimated due to the lack of demographic information in that study to control false positive regions with low genetic diversity (Bitocchi et al., 2017). Additionally, although Schmutz et al. (2014) reported that more than 90% of the genes related to domestication in dry beans were gene pool specific, most of them are physically close located, and domestication candidate genes for both gene pools were located within 22 out of 34 MQTL-YC supported by QTL identified in both gene pools, which give additional support to these regions in the control of domesticated traits as seed yield components (**Supplemental Table 2.11**).

Comparative genomic analysis

Comparative genomics is a useful approach to identify orthologous loci that have been reported in other species as causative loci. In this study, we compared the genome of common bean with soybean, pea, and rice genomes to identify orthologous MQTL (ortho-MQTL) across species. Since the genomic regions of 49 MQTL-YC were reported with signatures of selection (Schmutz et al., 2014; Wu et al., 2020), it was reasonable to hypothesize that similar legumes as pea and soybean could present orthoMQTL regions that control seed yield.

To assess the potential of comparative genomics to identify candidate genes, we chose MQTL-YC3.6 that presented synteny blocks within MQTL reported on soybean, pea, and rice (**Supplemental Table 2.9**). Kelly, 2018 suggested the importance of loci in chromosome PV03 related to yield in previous studies, although the loci were not well defined due to the availability of analytical tools when prior studies were reported. The gene Phvul.003G252400 located at 49.2

Mb has orthologous gene in soybean (Glyma.16G141100) and rice (LOC_Os02g45054) in the orthoMQTL regions. Phvul.003G252400 is a STK protein, and STK genes have been associated in rice with flowering (Deng et al., 2017) and seed weight (Hu et al., 2012), and in maize the overexpression of a STK protein significantly increased grain yield (Jia et al., 2020), which makes this gene a strong candidate gene in dry bean. A paralogous of Phvul.003G252400 were found in MQTL-YC7.5 (Phvul.007G174900 at 29.3 Mb). Interestingly, QTN associated with YDSD were closely located to both genes (Cichy et al., 2014; Resende et al., 2018), and they can be considered as candidate genes for seed yield.

Moreover, MQTL-YC1.8 showed synteny with a region related with SW in soybeans. This MQTL-YC was related with SDPD in four populations (Berny Mier Y Teran et al., 2019; Wu et al., 2020a; Nkhata et al., 2021; Diaz et al., 2022), and due to the relationship between SW and SDPD, the orthoMQTL could be controlled by orthologous genes. The gene Phvul.001G256200 (50.9 Mb) is an ABC transporter orthologous to Glyma.18g012700 in soybean. ABC transporters have been related with lodging resistance in soybean (Shook et al., 2021). Lodging has been negatively related with stem diameter (wider stem diameter better lodging) (Soltani et al., 2016), and bigger organs are frequently associated with larger seeds (Milla and Matesanz, 2017). Furthermore, a response regulator receiver gene (Glyma.09 g040000) has been related with SW and YDSD in soybean (Assefa et al., 2019), and an orthologous in dry bean (Phvul.001G258000) was presented in MQTL-YC1.8. The Phvul.001G256200 and Phvul.001G258000 presented closely located QTN related to SDPD (Wu et al., 2020b; Nkhata et al., 2021). These findings suggest that employing QTN, and comparative genomics could be useful to identify candidate genes in MQTL-YC regions.

Conclusions

In this study, we described a genome-wide landscape for the most consistent genomic regions associated with seed yield components through QTL and GWAS identified over the last two decades in common bean. All QTL and QTN were positioned on genome reference v2.1, which overcomes the limitation to compare results from older studies. To the best of our knowledge, this is the first MQTL for seed yield components in dry bean, and the most consistent regions can be used for candidate gene mining studies.

The robust MQTL-YC supported by at least five populations reported in this study should be considered for dry bean breeding to increase seed yield. The MQTL-YC regions could be used to

develop markers for marker-assisted selection and also into genomic selection models to assess their potential as fixed variables to increase the prediction accuracies (Izquierdo et al., 2018; McGowan et al., 2021). Functional genomics can be useful to identify candidate genes and to understand the interaction among genes in these genomic regions and their effect on seed yield. Furthermore, we found that the integration of QTL, GWAS and genomic information available from other legumes and related crops could help to improve our understanding of the genetic architecture of seed yield in dry bean.

BIBLIOGRAPHY

- Adams, M. W. (1967). Bases of yield component compensation in crop plants with special reference to the field bean, *Phaseolus vulgaris*. *Crop Science*, 7, 505–510.
- Assefa, T., Beebe, S. E., Rao, I. M., Cuasquer, J. B., Duque, M. C., Rivera, M., Battisti, A., & Lucchin, M. (2013). Pod harvest index as a selection criterion to improve drought resistance in white pea bean. *Field Crops Research*, 148, 24–33. <https://doi.org/10.1016/j.fcr.2013.04.008>
- Assefa, T., Otyama, P. I., Brown, A. V., Kalberer, S. R., Kulkarni, R. S., & Cannon, S. B. (2019). Genome-wide associations and epistatic interactions for internode number, plant height, seed weight and seed yield in soybean. *BMC Genomics*, 20(1), 1–12. <https://doi.org/10.1186/s12864-019-5907-7>
- Beebe, S. E., Rao, I. M., Blair, M. W., & Acosta-Gallegos, J. A. (2013). Phenotyping common beans for adaptation to drought. *Frontiers in Physiology*, 4 MAR(March), 1–20. <https://doi.org/10.3389/fphys.2013.00035>
- Beebe, S. E., Rao, I. M., Cajiao, C., & Grajales, M. (2008). Selection for drought resistance in common bean also improves yield in phosphorus limited and favorable environments. *Crop Science*, 48(2), 582–592. <https://doi.org/10.2135/cropsci2007.07.0404>
- Benjamin, D. J., Cesarini, D., Chabris, C. F., Glaeser, E. L., Laibson, D. I., Gunason, V., Harris, T. B., Launer, L. J., Purcell, S., Smith, A. V., Johannesson, M., Magnusson, P. K. E., Beauchamp, J. P., Christakis, N. A., Atwood, C. S., Hebert, B., Freese, J., Hauser, R. M., Hauser, T. S., ... Lichtenstein, P. (2012). The promises and pitfalls of geno-economics. *Annual Review of Economics*, 4, 627–662. <https://doi.org/10.1146/annurev-economics-080511-110939>
- Bernardo, R. (2008). Molecular markers and selection for complex traits in plants: Learning from the last 20 years. *Crop Science*, 48(5), 1649–1664. <https://doi.org/10.2135/cropsci2008.03.0131>
- Berny Mier Y Teran, J. C., Konzen, E. R., Palkovic, A., Tsai, S. M., Rao, I. M., Beebe, S., & Gepts, P. (2019). Effect of drought stress on the genetic architecture of photosynthate allocation and remobilization in pods of common bean (*Phaseolus vulgaris* L.), a key species for food security. *BMC Plant Biology*, 19(1), 1–15. <https://doi.org/10.1186/s12870-019-1774-2>
- Bilgrami, S., Liu, L., Farokhzadeh, S., Sobhani, A., Darzi, H., Nasiri, N., & Darwish, I. (2022). Industrial Crops & Products Meta-analysis of QTLs controlling seed quality traits based on QTL alignment in *Brassica napus*. *Industrial Crops & Products*, 176(August 2021), 114307. <https://doi.org/10.1016/j.indcrop.2021.114307>
- Bitocchi, E., Rau, D., Benazzo, A., Bellucci, E., Goretti, D., Biagetti, E., Panziera, A., Laidò, G., Rodriguez, M., Gioia, T., Attene, G., McClean, P., Lee, R. K., Jackson, S. A., Bertorelle, G., & Papa, R. (2017). High level of nonsynonymous changes in common bean suggests that selection under domestication increased functional diversity at target traits. *Frontiers in Plant Science*, 7(January), 1–15. <https://doi.org/10.3389/fpls.2016.02005>

- Blair, M. W., Iriarte, G., & Beebe, S. (2006). QTL analysis of yield traits in an advanced backcross population derived from a cultivated Andean x wild common bean (*Phaseolus vulgaris* L.) cross. *Theoretical and Applied Genetics*, 112(6), 1149–1163. <https://doi.org/10.1007/s00122-006-0217-2>
- Blair, M. W., & Izquierdo, P. (2012). Use of the advanced backcross-QTL method to transfer seed mineral accumulation nutrition traits from wild to Andean cultivated common beans. *TAG. Theoretical and Applied Genetics. Theoretische Und Angewandte Genetik*, 125(5), 1015–1031. <https://doi.org/10.1007/s00122-012-1891-x>
- Bolser, D., Staines, D., Pritchard, E., & Kersey, P. (2016). Ensembl Plants: Integrating Tools for Visualizing, Mining, and Analyzing Plant Genomics Data. In *Plant Bioinformatics Methods and Protocols* (Vol. 1374, pp. 115–140). <https://doi.org/10.1007/978-1-4939-3167-5>
- Brachi, B., Faure, N., Horton, M., Flahauw, E., Vazquez, A., Nordborg, M., Bergelson, J., Cuguen, J., & Roux, F. (2010). Linkage and association mapping of *Arabidopsis thaliana* flowering time in nature. *PLoS Genetics*, 6(5), 40. <https://doi.org/10.1371/journal.pgen.1000940>
- Brown, A. V., Grant, D., & Nelson, R. T. (2021). Using crop databases to explore phenotypes: From qtl to candidate genes. *Plants*, 10(11), 1–10. <https://doi.org/10.3390/plants10112494>
- Cano-Gamez, E., & Trynka, G. (2020). From GWAS to Function: Using Functional Genomics to Identify the Mechanisms Underlying Complex Diseases. *Frontiers in Genetics*, 11(May), 1–21. <https://doi.org/10.3389/fgene.2020.00424>
- Chebib, J., & Guillaume, F. (2021). Pleiotropy or linkage? Their relative contributions to the genetic correlation of quantitative traits and detection by multitrait GWA studies. *Genetics*, 219(4). <https://doi.org/10.1093/GENETICS/IYAB159>
- Checa, O. E., & Blair, M. W. (2012). Inheritance of yield-related traits in climbing beans (*Phaseolus vulgaris* L.). *Crop Science*, 52(5), 1998–2013. <https://doi.org/10.2135/cropsci2011.07.0368>
- Cichy, K. A., Fernandez, A., Kilian, A., Kelly, J. D., Galeano, C. H., Shaw, S., Brick, M., Hodgkinson, D., & Troxtell, E. (2014). QTL analysis of canning quality and color retention in black beans (*Phaseolus vulgaris* L.). *Molecular Breeding*, 33(1), 139–154. <https://doi.org/10.1007/s11032-013-9940-y>
- Dash, S., Campbell, J. D., Cannon, E. K. S., Cleary, A. M., Huang, W., Kalberer, S. R., Karingula, V., Rice, A. G., Singh, J., Umale, P. E., Weeks, N. T., Wilkey, A. P., Farmer, A. D., & Cannon, S. B. (2016). Legume information system (LegumeInfo.org): A key component of a set of federated data resources for the legume family. *Nucleic Acids Research*, 44(D1), D1181–D1188. <https://doi.org/10.1093/nar/gkv1159>
- Deng, L., Li, L., Zhang, S., Shen, J., Li, S., Hu, S., Peng, Q., Xiao, J., & Wu, C. (2017). Suppressor of *rid1* (*SID1*) shares common targets with *RID1* on florigen genes to initiate floral transition in rice. *PLoS Genetics*, 13(2), 1–24. <https://doi.org/10.1371/journal.pgen.1006642>
- Diaz, L. M., Ricaurte, J., Tovar, E., Cajiao, C., Terán, H., Grajales, M., Polanía, J., Rao, I., Beebe, S., & Raatz, B. (2018). QTL analyses for tolerance to abiotic stresses in a common bean

- (*Phaseolus vulgaris* L.) population. PLoS ONE, 13(8), 1–26. <https://doi.org/10.1371/journal.pone.0202342>
- Diaz, S., Ariza-Suarez, D., Izquierdo, P., Lobaton, J. D., de la Hoz, J. F., Acevedo, F., Duitama, J., Guerrero, A. F., Cajiao, C., Mayor, V., Beebe, S. E., & Raatz, B. (2020). Genetic mapping for agronomic traits in a MAGIC population of common bean (*Phaseolus vulgaris* L.) under drought conditions. BMC Genomics, 21(1), 1–20. <https://doi.org/10.1186/s12864-020-07213-6>
- Diaz, S., Polania, J., Ariza-Suarez, D., Cajiao, C., Grajales, M., & Raatz, B. (2022). Genetic Correlation Between Fe and Zn Biofortification and Yield Components in a Common Bean (*Phaseolus vulgaris* L.). Frontiers in Plant Science, 12(January), 1–13. <https://doi.org/10.3389/fpls.2021.739033>
- Didinger, C., Foster, M., Bunning, M., & Thompson, H. (2022). Nutrition and Human Health Benefits of Dry Beans and other Pulses. Dry Beans and Pulses Production, Processing and Nutrition, 481–504. <https://doi.org/10.1002/9781118448298.ch14>
- Doebley, J. (2006). Unfallen grains: How ancient farmers turned weeds into crops. Science, 312(5778), 1318–1319. <https://doi.org/10.1126/science.1128836>
- Evangelou, E., & Ioannidis, J. P. A. (2013). Meta-analysis Meta-analysis methods for genome-wide association studies and beyond. Nature Publishing Group, 14, 379. <https://doi.org/10.1038/nrg3472>
- Geravandi, M., Cheghamirza, K., Farshadfar, E., & Gepts, P. (2020). QTL analysis of seed size and yield-related traits in an inter-genepool population of common bean (*Phaseolus vulgaris* L.). Scientia Horticulturae, 274. <https://doi.org/10.1016/j.scienta.2020.109678>
- Goffinet, B., & Gerber, S. (2000). Quantitative trait loci: a meta-analysis. Genetics, 155(1), 463–473.
- González, A. M., Yuste-Lisbona, F. J., Weller, J., Vander Schoor, J. K., Lozano, R., & Santalla, M. (2021). Characterization of QTL and Environmental Interactions Controlling Flowering Time in Andean Common Bean (*Phaseolus vulgaris* L.). Frontiers in Plant Science, 11(January), 1–16. <https://doi.org/10.3389/fpls.2020.599462>
- Gramzow, L., & Theissen, G. (2010). A hitchhiker’s guide to the MADS world of plants. Genome Biology, 11(6). <https://doi.org/10.1186/gb-2010-11-6-214>
- Guo, B., Sleper, D. A., Lu, P., Shannon, J. G., Nguyen, H. T., & Arelli, P. R. (2006a). QTLs associated with resistance to soybean cyst nematode in soybean: Meta-analysis of QTL locations. Crop Science, 46(2), 595–602. <https://doi.org/10.2135/cropsci2005.04-0036-2>
- Guo, B., Sleper, D. A., Lu, P., Shannon, J. G., Nguyen, H. T., & Arelli, P. R. (2006b). QTLs Associated with Resistance to Soybean Cyst Nematode in Soybean: Meta-Analysis of QTL Locations. Crop Science, 46(2), 595–602. <https://doi.org/10.2135/cropsci2005.04-0036-2>
- Hageman, A., & Volkenburgh, E. (2021). Sink Strength Maintenance Underlies Drought Tolerance in Common Bean. Plants, 10(3:489), 1–12. <https://doi.org/10.2135/1980.hybridizationofcrops.c17>

- Heilig, J. A., Beaver, J. S., Wright, E. M., Song, Q., & Kelly, J. D. (2017). QTL analysis of symbiotic nitrogen fixation in a black bean population. *Crop Science*, 57(1), 118–129. <https://doi.org/10.2135/cropsci2016.05.0348>
- Hnatuszko-Konka, K., Kowalczyk, T., Gerszberg, A., Wiktorek-Smagur, A., & Kononowicz, A. K. (2014). *Phaseolus vulgaris* - Recalcitrant potential. *Biotechnology Advances*, 32(7), 1205–1215. <https://doi.org/10.1016/j.biotechadv.2014.06.001>
- Hu, Z., He, H., Zhang, S., Sun, F., Xin, X., Wang, W., Qian, X., Yang, J., & Luo, X. (2012). A Kelch Motif-Containing Serine/Threonine Protein Phosphatase Determines the Large Grain QTL Trait in Rice. *Journal of Integrative Plant Biology*, 54(12), 979–990. <https://doi.org/10.1111/jipb.12008>
- Hummel, M., Hallahan, B. F., Brychkova, G., Ramirez-Villegas, J., Guwela, V., Chataika, B., Curley, E., McKeown, P. C., Morrison, L., Talsma, E. F., Beebe, S., Jarvis, A., Chirwa, R., & Spillane, C. (2018). Reduction in nutritional quality and growing area suitability of common bean under climate change induced drought stress in Africa. *Scientific Reports*, 8(1), 1–11. <https://doi.org/10.1038/s41598-018-33952-4>
- Ibrahim, A. K., Zhang, L., Niyitanga, S., Afzal, M. Z., Xu, Y., Zhang, L., Zhang, L., & Qi, J. (2020). Principles and approaches of association mapping in plant breeding. *Tropical Plant Biology*, 13(3), 212–224. <https://doi.org/10.1007/s12042-020-09261-4>
- Izquierdo, P., Astudillo, C., Blair, M. W., Iqbal, A. M., Raatz, B., & Cichy, K. A. (2018). Meta-QTL analysis of seed iron and zinc concentration and content in common bean (*Phaseolus vulgaris* L.). *Theoretical and Applied Genetics*, 131(8), 1645–1658. <https://doi.org/10.1007/s00122-018-3104-8>
- Jensen, E. S., Peoples, M. B., Boddey, R. M., Gresshoff, P. M., Henrik, H. N., Alves, B. J. R., & Morrison, M. J. (2012). Legumes for mitigation of climate change and the provision of feedstock for biofuels and biorefineries. A review. In *Agronomy for Sustainable Development* (Vol. 32, Issue 2). <https://doi.org/10.1007/s13593-011-0056-7>
- Jeuffroy, M. H., Baranger, E., Carrouée, B., De Chezelles, E., Gosme, M., Hénault, C., Schneider, A., & Cellier, P. (2013). Nitrous oxide emissions from crop rotations including wheat, oilseed rape and dry peas. *Biogeosciences*, 10(3), 1787–1797. <https://doi.org/10.5194/bg-10-1787-2013>
- Jia, H., Li, M., Li, W., Liu, L., Jian, Y., Yang, Z., Shen, X., Ning, Q., Du, Y., Zhao, R., Jackson, D., Yang, X., & Zhang, Z. (2020). A serine/threonine protein kinase encoding gene *KERNEL NUMBER PER ROW6* regulates maize grain yield. *Nature Communications*, 11(1). <https://doi.org/10.1038/s41467-020-14746-7>
- Kamfwa, K., Cichy, K. A., & Kelly, J. D. (2015). Genome-wide association analysis of symbiotic nitrogen fixation in common bean. *Theoretical and Applied Genetics*, 128(10), 1999–2017. <https://doi.org/10.1007/s00122-015-2562-5>
- Keller, B., Ariza-Suarez, D., de la Hoz, J., Aparicio, J. S., Portilla-Benavides, A. E., Buendia, H. F., Mayor, V. M., Studer, B., & Raatz, B. (2020). Genomic Prediction of Agronomic Traits in Common Bean (*Phaseolus vulgaris* L.) Under Environmental Stress. *Frontiers in Plant*

Science, 11, 1001. <https://doi.org/10.3389/fpls.2020.01001>

- Kelly, J. D. (2018). Developing improved varieties of common bean. In: Achieving sustainable cultivation of grain legumes (S. Sivasankar, D. Bergvinson, P. Gaur, S. Agrawal, S. Beebe, & M. Tamò (eds.); 1st ed.). Burleigh Dodds Science Publishing.
- Khahani, B., Tavakol, E., Shariati, V., & Rossini, L. (2021). Meta-QTL and ortho-MQTL analyses identified genomic regions controlling rice yield, yield-related traits and root architecture under water deficit conditions. *Scientific Reports* |, 11, 6942. <https://doi.org/10.1038/s41598-021-86259-2>
- Klein, A., Houtin, H., Rond-coissieux, C., Naudet-Huart, M., Touratier, M., Marget, P., & Burstin, J. (2020). Meta-analysis of QTL reveals the genetic control of yield-related traits and seed protein content in pea. *Scientific RepoRtS* |, 10, 15925. <https://doi.org/10.1038/s41598-020-72548-9>
- Korte, A., & Farlow, A. (2013). The advantages and limitations of trait analysis with GWAS: a review. *Plant Methods*, 9(1), 29. <https://doi.org/10.1186/1746-4811-9-29>
- Kreplak, J., Madoui, M. A., Cápál, P., Novák, P., Labadie, K., Aubert, G., Bayer, P. E., Gali, K. K., Syme, R. A., Main, D., Klein, A., Bérard, A., Vrbová, I., Fournier, C., d'Agata, L., Belser, C., Berrabah, W., Toegelová, H., Milec, Z., ... Burstin, J. (2019). A reference genome for pea provides insight into legume genome evolution. *Nature Genetics*, 51(9), 1411–1422. <https://doi.org/10.1038/s41588-019-0480-1>
- Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S. J., & Marra, M. A. (2009). Circos: An information aesthetic for comparative genomics. *Genome Research*, 19(9), 1639–1645. <https://doi.org/10.1101/gr.092759.109>
- Li, H., Bradbury, P., Ersoz, E., Buckler, E. S., & Wang, J. (2011). Joint QTL linkage mapping for multiple-cross mating design sharing one common parent. *PLoS ONE*, 6(3). <https://doi.org/10.1371/journal.pone.0017573>
- MacQueen, A. H., Khoury, C. K., Miklas, P., McClean, P. E., Osorno, J. M., Runck, B. C., White, J. W., Kantar, M., & Ewing, P. M. (2021). Local to continental-scale variation in fitness and heritability in common bean (*Phaseolus vulgaris*) . *Crop Science*, 1–34. <https://doi.org/10.1002/csc2.20694>
- McGowan, M., Wang, J., Dong, H., Liu, X., Jia, Y., Wang, X., Iwata, H., Li, Y., Lipka, A. E., & Zhang, Z. (2021). Ideas in Genomic Selection with the Potential to Transform Plant Molecular Breeding. *Plant Breeding Reviews*, 45, 273–319. <https://doi.org/10.1002/9781119828235.ch7>
- Medendorp, J., Deyoung, D., Deepa, G., Duckworth, R., & Pittendrigh, B. (2022). 21 A Systems Perspective of the Role of Dry Beans and Pulses in the Future of Global Food Security: Opportunities and Challenges.
- Milla, R., & Matesanz, S. (2017). Growing larger with domestication: a matter of physiology, morphology or allocation? *Plant Biology*, 19(3), 475–483. <https://doi.org/10.1111/plb.12545>
- Moghaddam, S. M., Mamidi, S., Osorno, J. M., Lee, R., Brick, M., Kelly, J., Miklas, P., Urrea, C.,

- Song, Q., Cregan, P., Grimwood, J., Schmutz, J., & McClean, P. E. (2016). Genome-Wide Association Study Identifies Candidate Loci Underlying Agronomic Traits in a Middle American Diversity Panel of Common Bean. *The Plant Genome*, 9(november), 1–21. <https://doi.org/10.3835/plantgenome2016.02.0012>
- Mukeshimana, G., Butare, L., Cregan, P. B., Blair, M. W., & Kelly, J. D. (2014). Quantitative Trait Loci Associated with Drought Tolerance in Common Bean. *Crop Science*, 54(3), 923. <https://doi.org/10.2135/cropsci2013.06.0427>
- Nabateregga, M., Mukankusi, C., Raatz, B., Edema, R., Nkalubo, S., & Alladassi, B. M. E. (2019). Quantitative trait loci (QTL) mapping for intermittent drought tolerance in BRB 191 × SEQ 1027 Andean Intra-gene cross recombinant inbred line population of common bean (*Phaseolus vulgaris* L.). 18(21), 452–461. <https://doi.org/10.5897/AJB2019.16768>
- Nkhata, W., Hussein, S., Rob, M., Rowland, C., Tenyson, M., Isack, M., & Admire, S. (2021). Genome-wide association analysis of bean fly resistance and agro-morphological traits in common bean. 1–24. <https://doi.org/10.1371/journal.pone.0250729>
- Ouellette, L. A., Reid, R. W., Blanchard, S. G., & Brouwer, C. R. (2018). LinkageMapView-rendering high-resolution linkage and QTL maps. *Bioinformatics*, 34(2), 306–307. <https://doi.org/10.1093/bioinformatics/btx576>
- Ouyang, S., Zhu, W., Hamilton, J., Lin, H., Campbell, M., Childs, K., Thibaud-Nissen, F., Malek, R. L., Lee, Y., Zheng, L., Orvis, J., Haas, B., Wortman, J., & Buell, R. C. (2007). The TIGR Rice Genome Annotation Resource: Improvements and new features. *Nucleic Acids Research*, 35(SUPPL. 1), 8–11. <https://doi.org/10.1093/nar/gkl976>
- Polania, J., Poschenrieder, C., Beebe, S., & Rao, I. M. (2016). Effective Use of Water and Increased Dry Matter Partitioned to Grain Contribute to Yield of Common Bean Improved for Drought Resistance. *Frontiers in Plant Science*, 7(660), 1–10. <https://doi.org/10.3389/fpls.2016.00660>
- Rau, D., Murgia, M. L., Rodriguez, M., Bitocchi, E., Bellucci, E., Fois, D., Albani, D., Nanni, L., Gioia, T., Santo, D., Marcolungo, L., Delledonne, M., Attene, G., & Papa, R. (2019). Genomic dissection of pod shattering in common bean: mutations at non-orthologous loci at the basis of convergent phenotypic evolution under domestication of leguminous species. *Plant Journal*, 97(4), 693–714. <https://doi.org/10.1111/tpj.14155>
- Resende, R. T., de Resende, M. D. V., Azevedo, C. F., Silva, F. F. e., Melo, L. C., Pereira, H. S., Souza, T. L. P. O., Valdissier, P. A. M. R., Brondani, C., & Vianello, R. P. (2018). Genome-wide association and Regional Heritability Mapping of plant architecture, lodging and productivity in phaseolus vulgaris. *G3: Genes, Genomes, Genetics*, 8(8), 2841–2854. <https://doi.org/10.1534/g3.118.200493>
- Schmutz, J., Cannon, S. B., Schlueter, J., Ma, J., Mitros, T., Nelson, W., Hyten, D. L., Song, Q., Thelen, J. J., Cheng, J., Xu, D., Hellsten, U., May, G. D., Yu, Y., Sakurai, T., Umezawa, T., Bhattacharyya, M. K., Sandhu, D., Valliyodan, B., ... Jackson, S. A. (2010). Genome sequence of the palaeopolyploid soybean. *Nature*, 465(7294), 120–120. <https://doi.org/10.1038/nature08957>

- Schmutz, J., McClean, P. E., Mamidi, S., Wu, G. A., Cannon, S. B., Grimwood, J., Jenkins, J., Shu, S., Song, Q., Chavarro, C., Torres-Torres, M., Geffroy, V., Moghaddam, S. M., Gao, D., Abernathy, B., Barry, K., Blair, M., Brick, M. a, Chovatia, M., ... Jackson, S. a. (2014). A reference genome for common bean and genome-wide analysis of dual domestications. *Nature Genetics*, 46(November 2013), 707–713. <https://doi.org/10.1038/ng.3008>
- Shariatipour, N., Heidari, B., Ravi, S., & Stevanato, P. (2021). Genomic analysis of ionome-related QTLs in *Arabidopsis thaliana*. *Scientific Reports*, 11(1), 1–14. <https://doi.org/10.1038/s41598-021-98592-7>
- Shook, J. M., Zhang, J., Jones, S. E., Singh, A., Diers, B. W., & Singh, A. K. (2021). Meta-GWAS for quantitative trait loci identification in soybean. <https://doi.org/10.1093/g3journal/jkab117>
- Siddiq, M., & Uebersax, M. A. (2022). Overview , Production and Postharvest Technologies Dry Beans and Pulses Production and Consumption — An Overview. 3–28.
- Soltani, A., Bello, M., Mndolwa, E., Schroder, S., Moghaddam, S. M., Osorno, J. M., Miklas, P. N., Mcclean, P. E., Soltani, A., Schroder, S., Moghaddam, S. M., & Osorno, J. M. (2016). Targeted Analysis of Dry Bean Growth Habit : Interrelationship among Architectural , Phenological , and Yield Components. 3015(december), 3005–3015. <https://doi.org/10.2135/cropsci2016.02.0119>
- Song, Q., Jia, G., Hyten, D. L., Jenkins, J., Hwang, E.-Y., Schroeder, S. G., Osorno, J. M., Schmutz, J., Jackson, S. A., McClean, P. E., & Cregan, P. B. (2015). SNP Assay Development for Linkage Map Construction, Anchoring Whole-Genome Sequence, and Other Genetic and Genomic Applications in Common Bean. *Genes|Genomes|Genetics*, 5(11), 2285–2290. <https://doi.org/10.1534/g3.115.020594>
- Soriano, J. M., & Alvaro, F. (2019). Discovering consensus genomic regions in wheat for root-related traits by QTL meta-analysis. *Scientific Reports*, 9(1), 10537. <https://doi.org/10.1038/s41598-019-47038-2>
- Sosnowski, O., Charcosset, A., & Joets, J. (2012a). Biomecator V3: An upgrade of genetic map compilation and quantitative trait loci meta-analysis algorithms. *Bioinformatics*, 28(15), 2082–2083. <https://doi.org/10.1093/bioinformatics/bts313>
- Sosnowski, O., Charcosset, A., & Joets, J. (2012b). Biomecator V3: An upgrade of genetic map compilation and quantitative trait loci meta-analysis algorithms. *Bioinformatics*, 28(15), 2082–2083. <https://doi.org/10.1093/bioinformatics/bts313>
- Trapp, J. J., Urrea, C. A., Cregan, P. B., & Miklas, P. N. (2015). Quantitative trait loci for yield under multiple stress and drought conditions in a dry bean population. *Crop Science*, 55(4), 1596–1607. <https://doi.org/10.2135/cropsci2014.11.0792>
- Vandemark, G. J., Brick, M. A., Osorno, J. M., Kelly, J. D., & Urrea, C. A. (2015). Edible grain legumes. Yield Gains in Major U.S. Field Crops, 87–123. <https://doi.org/10.2135/cssaspepub33.c5>
- Veyrieras, J.-B., Goffinet, B., & Charcosset, A. (2007). MetaQTL: a package of new computational methods for the meta-analysis of QTL mapping experiments.

<https://doi.org/10.1186/1471-2105-8-49>

- Visscher, P. M., & Goddard, M. E. (2004). Prediction of the Confidence Interval of Quantitative Trait Loci Location. *Behavior Genetics*, 34(4).
- Wang, M., Li, W., Fang, C., Xu, F., Liu, Y., Wang, Z., Yang, R., Zhang, M., Liu, S., Lu, S., Lin, T., Tang, J., Wang, Y., Wang, H., Lin, H., Zhu, B., Chen, M., Kong, F., Liu, B., ... Tian, Z. (2018). Parallel selection on a dormancy gene during domestication of crops from multiple families. *Nature Genetics*, 50(10), 1435–1441. <https://doi.org/10.1038/s41588-018-0229-2>
- Weller, J. L., Vander Schoor, J. K., Perez-Wright, E. C., Hecht, V., González, A. M., Capel, C., Yuste-Lisbona, F. J., Lozano, R., & Santalla, M. (2019). Parallel origins of photoperiod adaptation following dual domestications of common bean. *Journal of Experimental Botany*, 70(4), 1209–1219. <https://doi.org/10.1093/jxb/ery455>
- Wickham, H. (2009). Elegant graphics for data analysis. *Springer*, 35(July), 211. <https://doi.org/10.1007/978-0-387-98141-3>
- Wright, D., & Kelly. (2011). Mapping QTL for seed yield and canning quality following processing of black bean (*Phaseolus vulgaris* L.). *Euphytica*, 179, 471–484. <https://doi.org/10.1007/s10681-011-0369-2>
- Wu, J., Wang, L., Fu, J., Chen, J., Wei, S., Zhang, S., Zhang, J., Tang, Y., Chen, M., Zhu, J., Lei, L., Geng, Q., Liu, C., Wu, L., Li, X., Wang, X., Wang, Q., Wang, Z., Xing, S., ... Wang, S. (2020a). Resequencing of 683 common bean genotypes identifies yield component trait associations across a north–south cline. *Nature Genetics*, 52(1), 118–125. <https://doi.org/10.1038/s41588-019-0546-0>
- Wu, J., Wang, L., Fu, J., Chen, J., Wei, S., Zhang, S., Zhang, J., Tang, Y., Chen, M., Zhu, J., Lei, L., Geng, Q., Liu, C., Wu, L., Li, X., Wang, X., Wang, Q., Wang, Z., Xing, S., ... Wang, S. (2020b). Resequencing of 683 common bean genotypes identifies yield component trait associations across a north–south cline. *Nature Genetics*, 52(1), 118–125. <https://doi.org/10.1038/s41588-019-0546-0>
- Zhou, X., Wang, J., Peng, C., Zhu, X., Yin, J., Li, W., He, M., Wang, J., Chern, M., Yuan, C., Wu, W., Ma, W., Qin, P., Ma, B., Wu, X., Li, S., Ronald, P., & Chen, X. (2016). Four receptor-like cytoplasmic kinases regulate development and immunity in rice. *Plant Cell and Environment*, 39(6), 1381–1392. <https://doi.org/10.1111/pce.12696>

APPENDIX

Table S2.1 Sequence and physical position of SSR associated with QTL (First 5 rows).

Marker	Sequence	Length	Chr	Start	End
BMc224	CCACTTTATAATTTCTACTACTTCTCTCTCTCTCTCTC TCTTACTCTCTGGGGGTCTGCATTCACTAGATACGGG GAGAACCGACACCACCGTGAATTGACCTTC	108	1	47,409,645	47,409,539
BMc294	GGTCGTGATGTCTCCATTTTTGTTTACCTTTTGTGCATA TATATATATATACATATATATATATATACATATATATAT ATATATATATGTGTGTGTGTGTATGTAAGTGCATGAT GGGAGGTA	124	7	32,100,636	32,100,758
BM172	ACTGTAGCTCAAACAGGGCACTGTAGCAGGTGTTGAC TTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG AGAGAGAGAGAGAGGAATCTCTCATGGCGGTATTGC TAAATATTGCTAGTTGT	126	2	45,866,746	45,866,616
BMc280	GAGGAAATATTGGNNGCAGAAATTGAAGAGAGAGA GAGAGAGAGAGAGAGAGAGAGAGAGAGGGTTGGATTG ATTTTCTTCTCCTTTTTCTAAGAGAATCTAATCTCAA TCCTCAACCTCCTTATTGCG	130	2	42,426,509	42,426,637
PVBR113	TGTGCATTCTTCTCCCATCTTCTCTCTCTCTTTCT CTCTCTCTCTCTCTCTCTCCAACACTACATAGTTTTATAC ACCACTGATCAAATCAAATCAAACCATTTAACTACTGT CAGACTTACCCTTTAC	135	11	2,276,514	2,276,647

Table S2.2 Sequence and physical position of SNPs associated with QTL and QTN (First row).

Paper	SNP	Chr	Position v1*	Position v2**	Sequence
Wu et al. 2020	Chr01_35934 2	1	359,342	229,264	AAGAAGATAAAATGCTCAGTCTATGAAC CTAATAAAATAGTTATAATATATATTTAC ATAAAAAATTAACCTTTTCTCATTAAATA TTATATGGAATAATAGTTTACCCCGGTC TATACGATAGGTGAAAGTTTGTATATCT GGTCGATCGACCGACACAACCAAATTCA TTATACTTAACAGAAAATAATGAAATT AACAGTAATCAAGTATTAGGTAATATA CTTTTAAATCATGATCTAAAACATTAATT AATCCAAATTAATAAAGCTTATGAAAA TAATATAAATAATTACAGATTTAAGGA TAAGGTACGTCGTTATTCTACACTAATA GCGTCGAACATCAAATACCAATCTCATT TAAACATCAGAGTGTCTTTTATCATTCAA GTTTGAAATTTACGAGAAGAAAGATAAT CATTCTTAGTTTAAATCTCATTGATATC AAATAATTTAGAAATTAGTTTACTGTAA TTGAAATTAATAAGTAAAAATGAAATA AACATTTAATTCTGTTTCTAAAAAAGAC AAATATCTTTGTGGTCCTTCAGCACTGA TGGCATGGCAACAATGTATGGTTGGT GGTGGTG

Table S2.3 QTL included in the MQTL analysis (First 20 rows).

Paper	Pop size	Cross type	TRAIT	QTL ID	Site	Year	Chr	QTL position Mb	QTL position cM	LOD	R2	Add	CI	Environment
Asfaw2012	97	RI0	PHI	As12_PHI8.1_Pa07D	Palmira (Co)	2007	8	12.90	60.86	2.20	0.08	0.57	21.0	Drought
Asfaw2012	97	RI0	PHI	As12_PHI1.1_Aw09N	Awassa (Et)	2009	1	32.41	50.08	3.24	0.12	1.52	14.0	Non-drought
Asfaw2012	97	RI0	YDSD	As12_YDSD1.1_Am09N	Amaro (Co)	2009	1	32.41	50.08	2.43	0.11	169.91	15.3	Non-drought
Asfaw2012	97	RI0	YDSD	As12_YDSD8.1_Pa07D	Palmira (Co)	2007	8	39.35	68.20	2.90	0.14	48.22	12.0	Drought
Asfaw2012	97	RI0	PHI	As12_PHI1.1_Pa07N	Palmira (Co)	2007	1	47.41	75.65	4.73	0.26	0.78	6.5	Non-drought
Asfaw2012	97	RI0	PHI	As12_PHI6.1_Pa07N	Palmira (Co)	2007	6	21.01	62.34	3.98	0.33	0.92	5.1	Non-drought
Asfaw2012	97	RI0	HI	As12_HI9.1_Aw09N	Awassa (Et)	2009	9	14.06	30.92	2.27	0.11	4.78	15.3	Non-drought
Blair2012	113	RI0	SW	Bla12_SW9.1_Pa07D	Palmira (Co)	2007	9	23.13	62.92	5.16	0.15	0.87	9.6	Drought
Blair2012	113	RI0	DPM	Bla12_DPM6.2_Pa06D	Palmira (Co)	2006	1	6.09	21.93	3.42	0.13	-0.77	11.1	Drought
Blair2012	113	RI0	DF	Bla12_DF5.1_Pa05N	Palmira (Co)	2005	3	11.67	54.87	3.09	0.12	0.42	12.0	Non-drought
Blair2012	113	RI0	SW	Bla12_SW6.9_Pa06N	Palmira (Co)	2006	6	8.54	28.67	5.27	0.22	1.02	6.6	Non-drought
Blair2012	113	RI0	SW	Bla12_SW6.5_Pa06D	Palmira (Co)	2006	6	8.54	28.67	3.86	0.16	0.92	9.0	Drought
Blair2012	113	RI0	YDSD	Bla12_YDSD3.3_Pa06N	Palmira (Co)	2006	5	37.08	81.02	3.27	0.14	155.68	10.3	Non-drought
Blair2012	113	RI0	SW	Bla12_SW6.16_Pa07N	Palmira (Co)	2007	6	19.57	51.65	8.06	0.30	1.28	4.8	Non-drought
Blair2012	113	RI0	SW	Bla12_SW6.8_Pa06D	Palmira (Co)	2006	6	19.57	51.65	5.48	0.21	1.05	6.9	Drought
Blair2012	113	RI0	SW	Bla12_SW6.13_Pa07D	Palmira (Co)	2007	6	19.57	51.65	6.18	0.17	0.95	8.5	Drought
Blair2012	113	RI0	DPM	Bla12_DPM6.3_Pa05N	Palmira (Co)	2005	6	21.01	62.34	4.40	0.50	0.97	2.9	Non-drought
Checa-Blair2012	84	RI0	YDSD	Ch12_YDSD3.1_PoN	Popayan (Co)	-	3	36.83	78.46	3.14	0.12	-5.17	16.8	Non-drought
Checa-Blair2012	84	RI0	SW	Ch12_SW6.2_PoN	Popayan (Co)	-	6	19.57	51.65	3.53	0.14	-2.75	13.9	Non-drought
Checa-Blair2012	84	RI0	DPM	Ch12_DPM9.2_PoN	Popayan (Co)	-	9	1.82	2.49	4.80	0.20	-3.30	9.7	Non-drought

Table S2.4 QTN included in the MQTL analysis (First 20 rows).

Paper	TRAIT	QTN unified	QTN ID	Site	Year	Chr	QTN position bp	QTN position cM	pvalue	R2	Notes	Env
Galeano2012	PDPL	PDPL2.1	Ga12_PDPL2.1_Pa09D	Palmira (Co)	2009	2	2,757,890	39.76	0.00	0.48	GLM	Drought
Galeano2012	PDPL	PDPL2.1	Ga12_PDPL2.1_Pa09N	Palmira (Co)	2009	2	2,757,890	39.76	0.00	0.46	GLM	Non-drought
Galeano2012	YDSD	YDSD2.1	Ga12_YDSD2.1_Pa09N	Palmira (Co)	2009	2	23,186,699	77.60	0.01	0.10	MLM	Non-drought
Galeano2012	YDSD	YDSD2.2	Ga12_YDSD2.2_Pa09N	Palmira (Co)	2009	2	23,681,150	81.94	0.01	0.12	MLM	Non-drought
Galeano2012	PDPL	PDPL9.1	Ga12_PDPL9.1_Pa09D	Palmira (Co)	2009	2	29,180,769	90.29	0.03	0.09	MLM	Drought
Galeano2012	PDPL	PDPL9.1	Ga12_PDPL9.1_Pa09N	Palmira (Co)	2009	2	29,180,769	90.29	0.05	0.08	MLM	Non-drought
Galeano2012	DPM	DPM2.1	Ga12_DPM2.1_Pa09D	Palmira (Co)	2009	2	45,148,809	139.69	0.00	0.19	MLM	Drought
Galeano2012	DPM	DPM2.1	Ga12_DPM2.1_Pa09N	Palmira (Co)	2009	2	45,148,809	139.69	0.00	0.24	MLM	Non-drought
Galeano2012	PDPL	PDPL2.1	Ga12_PDPL2.1_Pa09D	Palmira (Co)	2009	2	45,148,809	139.69	0.01	0.16	MLM	Drought
Galeano2012	PDPL	PDPL2.1	Ga12_PDPL2.1_Pa09N	Palmira (Co)	2009	2	45,148,809	139.69	0.01	0.15	MLM	Non-drought
Galeano2012	SDPD	SDPD2.1	Ga12_SDPD2.1_Pa09D	Palmira (Co)	2009	2	45,148,809	139.69	0.02	0.13	MLM	Drought
Galeano2012	YDSD	YDSD2.3	Ga12_YDSD2.3_Pa09D	Palmira (Co)	2009	2	45,148,809	139.69	0.04	0.12	MLM	Drought
Galeano2012	YDSD	YDSD2.3	Ga12_YDSD2.3_Pa09N	Palmira (Co)	2009	2	45,148,809	139.69	0.04	0.12	MLM	Non-drought
Galeano2012	PDPL	PDPL2.2	Ga12_PDPL2.2_Pa09D	Palmira (Co)	2009	2	45,866,746	150.57	0.00	0.38	GLM	Drought
Galeano2012	DF	DF3.1	Ga12_DF3.1_Pa09N	Palmira (Co)	2009	3	28,340,387	63.76	0.04	0.13	MLM	Non-drought
Galeano2012	DF	DF3.2	Ga12_DF3.2_Pa09D	Palmira (Co)	2009	3	39,901,377	83.58	0.00	0.14	MLM	Drought
Galeano2012	DF	DF3.2	Ga12_DF3.2_Pa09N	Palmira (Co)	2009	3	39,901,377	83.58	0.00	0.15	MLM	Non-drought
Galeano2012	DPM	DPM3.1	Ga12_DPM3.1_Pa09D	Palmira (Co)	2009	3	39,901,377	83.58	0.02	0.07	MLM	Drought
Galeano2012	SDPD	SDPD3.1	Ga12_SDPD3.1_Pa09N	Palmira (Co)	2009	3	39,901,377	83.58	0.02	0.08	MLM	Non-drought
Galeano2012	SW	SW3.1	Ga12_SW3.1_Pa09N	Palmira (Co)	2009	3	39,901,377	83.58	0.02	0.08	MLM	Non-drought

Table S2.5 QTL within MQTL-YC regions (MQTL-YC1.1).

Chr	MQTL	ID QTL	TRAIT	LOD	R2	Add	QTL Mb	QTL cM	Location	Latitude	Source	Gene-pool source	Environment
Pv01	MQTL-YC1.1	Ge20_HI1.1_Ke14N	HI	2.56	0.1	2.45	0.19	0.77	Kermanshah (Ir)	34.38	AND1007	A	Non-drought
		Di22_PHI1.1_Pa13D	PHI	3.16	0.1	2.1	0.90	3.73	Palmira (Co)	3.53	SCR16	M	Drought
		Sa18_SW1.1_Mo16N	SW	5.9	0.1	0.5	1.60	6.61	Morden (Ca)	49.19	BK004-001	M	Non-drought
		Si18_SW1.1_Co12N	SW	7	0.1	1.96	2.66	10.77	Coimbra (Bra)	-16.68	AND277	A	Non-drought
		Blb12_SW3.1_Da04N	SW	8.77	0.2	5.7	2.85	10.99	Darien (Co)	3.91	Cerinza	A	Non-drought
		Blb12_SW3.2_Pa99N	SW	6.06	0.1	4.9	2.85	10.99	Palmira (Co)	3.53	Cerinza	A	Non-drought
		Tr15_DF1.1_Mi11N	DF	5.1	0.2	0.5	3.22	12.47	Mitchell,NE (US)	41.94	Buster	M	Non-drought
		Tr15_DF1.1_Mi12N	DF	4.2	0.1	0.3	3.22	12.47	Mitchell,NE (US)	41.94	Buster	M	Non-drought
		Tr15_DF1.1_Ot11N	DF	5.1	0.2	0.8	3.22	12.47	Othello, WA (US)	46.83	Buster	M	Non-drought
		Tr15_DF1.1_Ot12N	DF	5.1	0.2	1.2	3.22	12.47	Othello, WA (US)	46.83	Buster	M	Non-drought
Tr15_SW1.1_Ot11D	SW	4	0.1	0.3	3.22	12.47	Othello, WA (US)	46.83	Buster	M	Drought		

Table S2.6 QTN within MQTL-YC regions (MQTL-YC1.1 and MQTL-YC1.2).

MQTL	ID QTL	TRAIT	Mb	cM	Location	Latitude	Genepool	Pop size	Env	Year	p-value	Study
MQTL-YC1.1	Wu20_DF1.1_Bi14N	DF	0.23	0.95	Bijie (Chi)	27.3	M - A	683	Non-drought	2014	0.00	Wuetal2020
	Di20_YDSD1.2_Pa14D	YDSD	0.25	1.05	Palmira (Co)	3.5	M	636	Drought	2014	0.00	Diaz2020
	Di20_DPM1.1_Pa14D	DPM	0.76	3.15	Palmira (Co)	3.5	M	636	Drought	2014	0.00	Diaz2020
	Wu20_DPM1.1_Bi14N	DPM	0.87	3.58	Bijie (Chi)	27.3	M - A	683	Non-drought	2014	0.00	Wuetal2020
	Nk21_YDSD1.1_JoN	YDSD	1.64	6.76	Chitedze-Mbawa (Ma)	-12.9	M - A	99	Non-drought	2018-2019	0.00	Nkhata2021
	Di20_PHI1.1_Pa13D	PHI	2.16	8.92	Palmira (Co)	3.5	M	636	Drought	2013	0.00	Diaz2020
	Wu20_DF1.2_Bi14N	DF	2.92	11.06	Bijie (Chi)	27.3	M - A	683	Non-drought	2014	0.00	Wuetal2020
MQTL-YC1.2	Mi21_PDPL6.1_JaN	PDPL	6.09	21.93	Jammu (In)	32.7	M - A	96	Non-drought	-	0.03	Mir2021
	Vi15_SW1.1_CeC	SW	6.15	22.09	Celaya (Mex)	20.5	M	282	Combined	2010-2011	0.00	Villordo2015
	Nk21_YDSD1.2_JoN	YDSD	6.47	22.99	Chitedze-Mbawa (Ma)	-12.9	M - A	99	Non-drought	2018-2019	0.00	Nkhata2021
	Vi15_DF1.1_CeC	DF	6.57	23.23	Celaya (Mex)	20.5	M	282	Combined	2010-2011	0.00	Villordo2015
	Wu20_SW1.1_Sn15N	SW	6.91	23.79	Sanya (Chi)	18.3	M - A	683	Non-drought	2015	0.00	Wuetal2020
	Mo16_DF1.1_JoN	DF	7.91	25.50	CO, NE (US)	40.7	M	280	Non-drought	-	0.00	Moghaddam2016
	Di20_DPM1.2_Pa13D	DPM	10.53	27.04	Palmira (Co)	3.5	M	636	Drought	2013	0.00	Diaz2020
	Di20_YDSD1.1_Pa14D	YDSD	11.22	27.04	Palmira (Co)	3.5	M	636	Drought	2014	0.00	Diaz2020
	Di20_YDSD1.1_Pa13D	YDSD	11.25	27.04	Palmira (Co)	3.5	M	636	Drought	2014	0.00	Diaz2020
	Di20_DPM1.2_Pa14D	DPM	11.56	29.69	Palmira (Co)	3.5	M	636	Drought	2014	0.00	Diaz2020
	Wu20_DF1.3_Sn16N	DF	13.56	29.82	Sanya (Chi)	18.3	M - A	683	Non-drought	2016	0.00	Wuetal2020
	Mo16_DF1.2_JoN	DF	13.75	29.84	CO, MI, ND (US)	43.5	M	280	Non-drought	-	0.00	Moghaddam2016
	Di20_DF1.2_Pa13D	DF	14.67	29.95	Palmira (Co)	3.5	M	636	Drought	2013	0.00	Diaz2020
	Nk21_PDPL1.1_JoN	PDPL	15.95	30.90	Chitedze-Mbawa (Ma)	-12.9	M - A	99	Non-drought	2018-2019	0.00	Nkhata2021
	Mo16_DF1.3_JoN	DF	16.88	30.90	MI, CO (US)	41.4	M	280	Non-drought	-	0.00	Moghaddam2016
	Di20_DF1.2_Pa14D	DF	17.63	30.90	Palmira (Co)	3.5	M	636	Drought	2014	0.00	Diaz2020

Table S2.7 Candidate genes (First 7 rows).

MQTL	Name	Chr	start	end	QTN	QTN Mb position	distance Mb QTN - Gene	Gene description	Reference
MQTL-YC1.1	Phvul.001G027400	Pv01	2.45	2.45	Di20_PHI1.1_Pa13D	2.16	0.29	PTHR10992:SF891 - STRIGOLACTONE ESTERASE D14 HOMOLOG-RELATED	Yao et al 2018
MQTL-YC1.2	Phvul.001G085200	Pv01	13.12	13.13	Wu20_DF1.3_Sn16N	13.56	0.43	PTHR36762:SF2 - A.THALIANA MRNA (ORF19) FROM CHROMOSOME III	Ciannamea et al 2007
MQTL-YC1.2	Phvul.001G087300	Pv01	13.55	13.55	Wu20_DF1.3_Sn16N	13.56	0.00	PTHR31717:SF3 - ZINC FINGER PROTEIN CONSTANS-LIKE 14-RELATED	Tiwari et al 2010
MQTL-YC1.2	Phvul.001G089900	Pv01	14.66	14.66	Di20_DF1.2_Pa13D	14.67	0.01	PF03514 - GRAS domain family (GRAS)	Tyler et al 2004
MQTL-YC1.2	Phvul.001G094300	Pv01	16.83	16.83	Mo16_DF1.3_JoN	16.88	0.04	K11650 - SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily D (SMARCD)	Jégu et al 2014
MQTL-YC1.5	Phvul.001G176200	Pv01	43.30	43.30	Nk21_DF1.1_JoN	42.92	0.38	PTHR10593:SF42 - ZINC FINGER PROTEIN JACKDAW , SERINE/THREONINE-PROTEIN KINASE RIO	Jia et al 2020
MQTL-YC1.6	Phvul.001G186400	Pv01	44.43	44.43	Wu20_DF1.4_Nn16N	44.79	0.36	PTHR11945:SF185 - AGAMOUS-LIKE MADS-BOX PROTEIN AGL62	Gramzow and Theissen 2010

Table S2.8 Robust MQTL-YC supported by more than five populations.

Chr	MQTL	Pos (cM)	CI (cM)	Pos (Mb)	# of	# of populations	Env	GP	Lat
					QTL-QTN	QTL-QTN			
Pv01	MQTL-YC1.1	10.4	2.4	2.4	12-7	7-3	D6, N13	B	B
	MQTL-YC1.2	27.4	0.6	11.4	11-16	4-6	D15, N10, C2	B	B
	MQTL-YC1.4	51.0	2.0	37.1	11-3	5-2	D7, N7	B	B
	MQTL-YC1.6	65.8	1.2	45.0	4-6	3-2	D1, N9	M	B
	MQTL-YC1.7	74.5	1.2	47.2	16-6	4-2	D4, N18	B	B
Pv02	MQTL-YC2.2	41.8	3.5	2.9	8-2	5-2	D2, N7, C1	B	B
	MQTL-YC2.5	93.5	3.7	30.7	6-12	3-4	D5, N13	B	B
Pv03	MQTL-YC3.3	67.5	3.2	31.4	5-5	4-5	D1, N9	B	B
	MQTL-YC3.4	82.3	2.0	39.6	9-11	5-4	D8, N12	B	B
	MQTL-YC3.5	88.4	2.2	41.7	10-1	6-1	D1, N9, C1	B	B
	MQTL-YC3.6	115.7	0.8	46.4	8-3	5-2	D3, N8	B	B
Pv04	MQTL-YC4.1	4.2	2.6	0.5	7-6	6-5	D4, N8, C1	B	B
	MQTL-YC4.2	48.5	3.0	12.6	7-5	3-3	D4, N8	B	B
	MQTL-YC4.3	77.9	7.7	41.2	3-7	2-3	D4, N6	B	B
	MQTL-YC4.5	127.8	1.0	45.4	6-6	4-2	D1, N10, C1	B	B
Pv05	MQTL-YC5.4	102.6	1.5	39.2	8-2	4-1	N8, C2	B	B
Pv06	MQTL-YC6.1	30.8	3.4	12.2	6-11	3-3	D10, N7	B	B
	MQTL-YC6.5	87.3	0.5	26.8	9-11	4-5	D7, N11, C2	B	B
Pv07	MQTL-YC7.1	15.4	4.8	1.9	8-3	4-2	D5, N6	M	B
	MQTL-YC7.5	63.9	0.3	27.0	31-20	5-8	D23, N22, C6	B	B
Pv08	MQTL-YC8.1	9.4	1.9	1.0	7-5	4-4	D6, N6	M	B
	MQTL-YC8.3	72.7	3.7	44.5	9-7	6-4	D6, N10	B	B
	MQTL-YC8.5	141.3	0.0	59.5	9-10	6-6	D2, N16, C1	B	B
Pv09	MQTL-YC9.2	31.8	6.2	14.3	4-3	3-2	D1, N6	B	B
Pv10	MQTL-YC10.1	9.5	4.7	2.7	5-4	3-3	D3, N5, C1	M	B
	MQTL-YC10.2	85.8	8.4	40.9	3-7	3-2	D3, N7	B	B
Pv11	MQTL-YC11.1	5.9	3.3	0.6	9-2	3-2	D5, N6	B	B
	MQTL-YC11.3	76.5	29.2	38.7	2-8	1-6	D3, N6, C1	B	B

Table S2.9 Ortho MQTL-YC between common bean, soybean, pea, and rice.

MQTL	start (Mb)	end (Mb)	N of QTL-QTN	Traits	Genepool	*MQTL soybean	MQTL pea
MQTL-YC1.1	0.2	3.2	12-7	DF, DPM, HI, PHI, SW, YDSD	B	SW_ms967	-
MQTL-YC1.2	5.5	20.2	11-16	DF, DPM, PDPL, SDPD, SW, YDSD	B	-	mQTL5.2
MQTL-YC1.4	31.2	38.9	11-3	DF, DPM, PHI, SW, YDSD	B	-	mQTL5.2
MQTL-YC1.7	47.0	48.1	16-6	DF, DPM, PHI, SW, YDSD	B	-	mQTL5.1
MQTL-YC1.8	51.0	51.3	5-0	PHI, SDPD	M	MD_il989, SW_il164	-
MQTL-YC2.1	1.4	1.9	5-0	PDPL, SDPD, SW	M	SW_meta	mQTL1.4, mQTL4.3
MQTL-YC2.2	2.8	3.9	8-2	DF, HI, SW, YDSD	B	MD_ms1999.01	-
MQTL-YC2.3	8.1	12.3	7-0	DPM, PHI, SW, YDSD	M	MD_il989	-
MQTL-YC2.4	17.4	22.2	5-0	DF	M	-	mQTL7.2
MQTL-YC2.6	39.3	42.3	5-2	DF, DPM, PHI, SW	M	DF_ms1999.01, YIELD_meta	-
MQTL-YC2.7	46.8	47.7	5-2	PHI, SW	M	DF_ms923	mQTL7.1
MQTL-YC3.1	0.2	2.0	3-1	DF, PHI, SW, YDSD	M	YIELD_il0102	mQTL2.1-2.2
MQTL-YC3.2	2.6	4.1	9-0	PDPL, SW, YDSD	B	MD_il989	-
MQTL-YC3.3	11.7	33.0	5-5	DF, DPM, PDPL, SW, YDSD	B	SW_meta	mQTL4.1, mQTL4.2
MQTL-YC3.4	35.4	40.0	9-11	DF, DPM, SDPD, SW, YDSD	B	SW_meta	mQTL4.1, mQTL4.2
MQTL-YC3.6	44.3	49.2	8-3	DF, HI, SW, YDSD	B	YIELD_2il81.2	mQTL4.3
MQTL-YC4.1	0.2	2.1	7-6	DF, DPM, PHI, SDPD, SW, YDSD	B	DF_ms923	-
MQTL-YC4.2	8.6	20.3	7-5	DF, PDPL, SW, YDSD	B	-	mQTL6.2
MQTL-YC4.4	42.8	43.6	4-0	PHI, SDPD	M	-	mQTL6.2
MQTL-YC4.5	45.2	46.3	6-6	DF, DPM, SDPD, YDSD	B	-	mQTL6.1
MQTL-YC5.1	0.4	1.8	2-4	DPM, PHI, SW	B	-	mQTL3.3
MQTL-YC5.2	4.6	7.2	13-0	DF, PHI, SDPD, YDSD	B	DF_sojams989	mQTL6.1
MQTL-YC5.3	10.6	16.1	5-0	SW	M	-	mQTL3.3
MQTL-YC5.4	38.7	39.5	8-2	DF, DPM, PHI, SDPD, SW, YDSD	B	MD_2mn81	mQTL3.4
MQTL-YC6.1	8.5	17.1	6-11	DF, DPM, HI, PDPL, SDPD, SW, YDSD	B	YIELD_ms923	-
MQTL-YC6.4	21.0	21.7	4-0	DPM, PHI, SW	B	-	mQTL5.1
MQTL-YC6.5	24.4	30.3	9-11	DF, DPM, PHI, SW, YDSD	B	-	mQTL6.3
MQTL-YC7.1	0.8	2.4	8-3	DF, DPM, SW, YDSD	M	DF_meta, MD_meta, YIELD_meta	-
MQTL-YC7.2	4.4	5.6	4-0	SDPD, SW, YDSD	B	-	mQTL2.3
MQTL-YC7.5	11.1	39.6	31-20	DF, DPM, PDPL, PHI, SDPD, SW, YDSD	B	DF_il989, MD_il989	mQTL2.1-2.2
MQTL-YC8.1	0.3	1.8	7-5	DF, DPM, PHI, SW, YDSD	M	-	mQTL5.2-5.3
MQTL-YC8.4	52.7	55.0	3-0	SDPD, SW, YDSD	M	-	mQTL1.2-1.3
MQTL-YC8.5	58.2	62.8	9-10	DF, DPM, PDPL, SDPD, SW, YDSD	B	-	mQTL1.1
MQTL-YC9.2	14.0	15.4	4-3	DF, HI, PHI, SW, YDSD	B	SW_il0102	mQTL3.1
MQTL-YC9.3	18.0	25.6	7-1	DF, DPM, SW	B	-	mQTL3.2
MQTL-YC9.4	26.4	28.2	3-1	DF, YDSD	B	-	mQTL3.2
MQTL-YC9.5	28.9	30.0	2-0	DF, DPM	A	DF_meta, MD_meta	-
MQTL-YC10.1	1.2	5.5	5-4	DPM, PDPL, PHI, SW, YDSD	M	-	mQTL4.5
MQTL-YC10.2	40.5	41.3	3-7	DPM, PDPL, SDPD, YDSD	B	DF_5il90	mQTL4.5
MQTL-YC11.1	0.4	1.3	9-2	DF, SW, YDSD	B	DF_ms923	mQTL7.3
MQTL-YC11.2	5.3	6.1	6-0	DF, SW	M	-	mQTL7.4
MQTL-YC11.3	10.1	45.4	2-8	DF, DPM, PDPL, SW, YDSD	B	DF_meta, MD_il989, YIELD_mn945	mQTL7.3-7.4
MQTL-YC11.4	49.4	51.8	3-2	DF, PHI, YDSD	M	YIELD_ms967	mQTL4.5

Table S2.10 Syntenic blocks between common bean, soybean, pea, and rice within MQTL-YC identified in common bean (First 10 rows).

MQTL	Chr	strt	end	Chr	strt	end	Ortho MQTL	Species
MQTL-YC1.2	1	12,009,948	13,845,018	LG5chr3	187,952,271	192,418,130	mQTL5.2	Pea
MQTL-YC1.2	1	18,472,836	20,540,720	LG5chr3	184,289,381	187,946,539	mQTL5.2	Pea
MQTL-YC1.4	1	29,767,067	31,046,003	LG5chr3	169,998,654	173,522,317	mQTL5.2	Pea
MQTL-YC1.4	1	33,325,226	39,198,785	LG5chr3	140,420,101	170,216,805	mQTL5.2	Pea
MQTL-YC1.7	1	44,093,189	47,900,606	LG5chr3	32,842,937	86,514,356	mQTL5.1	Pea
MQTL-YC1.7	1	47,929,577	48,150,380	LG5chr3	29,173,139	32,528,101	mQTL5.1	Pea
MQTL-YC2.1	2	58,574	3,383,145	LG1chr2	389,416,003	415,611,929	mQTL1.4	Pea
MQTL-YC2.1	2	1,443,439	1,736,027	LG4chr4	145,466,627	149,435,633	mQTL4.3	Pea
MQTL-YC2.4	2	16,216,312	18,008,439	LG7chr7	209,697,413	217,268,596	mQTL7.2	Pea
MQTL-YC2.7	2	44,080,455	47,787,960	LG7chr7	23,576,558	79,789,474	mQTL7.1	Pea

Table S2.11 List of candidate genes within MQTL-YC regions related to domestication by Schmutz et al 2014 and Wu et al 2020 (First 10 rows).

Name	Gene pool	Chr	Start	End	MQTL_YC	Description	Reference
Phvul.001G000800	Andean	1	35,239	37,775	MQTL-YC1.1	PTHR22893//PTHR22893:SF62 - NADH OXIDOREDUCTASE-RELATED // SUBFAMILY NOT NAMED (1 of 2)	Schmutz et al 2014
Phvul.001G001200	Andean	1	62,008	69,638	MQTL-YC1.1	PF01535//PF13041 - PPR repeat (PPR) // PPR repeat family (PPR_2) (1 of 197)	Schmutz et al 2014
Phvul.001G004500	Middle American	1	271,490	275,181	MQTL-YC1.1	PTHR22835//PTHR22835:SF227 - ZINC FINGER FYVE DOMAIN CONTAINING PROTEIN // SUBFAMILY NOT NAMED (1 of 4)	Schmutz et al 2014
Phvul.001G008300	Middle American	1	619,173	620,063	MQTL-YC1.1	PTHR31384:SF10 - AUXIN RESPONSE FACTOR 5 (1 of 2)	Schmutz et al 2014
Phvul.001G009400	Middle American	1	709,104	712,781	MQTL-YC1.1	PTHR10992//PTHR10992:SF866 - ALPHA/BETA HYDROLASE FOLD-CONTAINING PROTEIN // SUBFAMILY NOT NAMED (1 of 2)	Schmutz et al 2014
Phvul.001G012200	Andean	1	928,512	932,896	MQTL-YC1.1	K10572 - inositol-pentakisphosphate 2-kinase (IPPK) (1 of 2)	Schmutz et al 2014
Phvul.001G013100	Andean	1	1,000,437	1,003,967	MQTL-YC1.1	PTHR10584//PTHR10584:SF153 - SUGAR KINASE // SUBFAMILY NOT NAMED (1 of 2)	Schmutz et al 2014
Phvul.001G013700	Andean	1	1,038,062	1,040,704	MQTL-YC1.1	PF01535//PF13041//PF13812 - PPR repeat (PPR) // PPR repeat family (PPR_2) // Pentatricopeptide repeat domain (PPR_3) (1 of 46)	Schmutz et al 2014
Phvul.001G014200	Andean	1	1,064,026	1,069,515	MQTL-YC1.1	PTHR11746:SF89 - MAF-LIKE PROTEIN (1 of 2)	Schmutz et al 2014
Phvul.001G014300	Andean	1	1,073,244	1,074,353	MQTL-YC1.1	PTHR24078:SF177 - PROTEIN DNJ-23-RELATED (1 of 12)	Schmutz et al 2014

**CHAPTER 3: GWAS-ASSISTED AND MULTI-TRAIT GENOMIC PREDICTION FOR
IMPROVEMENT OF SEED YIELD AND CANNING QUALITY TRAITS IN A BLACK
BEAN BREEDING PANEL**

Abstract

A major end use of dry beans (*Phaseolus vulgaris* L) in the U.S. is the canning market. Canning quality is an important consideration in the dry bean breeding process. However, the need for specialized equipment and trained sensory panels required to effectively evaluate canning quality limit the capacity to incorporate it into many dry bean breeding programs. The integration of genomics and phenomics has the potential to increase selection accuracy and intensity in traits that are difficult to measure and/or have a complex genetic architecture. Genomic prediction (GP) has been shown to improve breeding progress for complex traits. Furthermore, the use of phenotypic correlated traits in GP using multi-trait models can boost prediction accuracies. In this study, we assessed the prediction accuracy of single-trait and multi-trait GP models and the use of significant markers identified through genome-wide association (GWAS) in a black bean advanced breeding lines over two breeding cycles. We identified that prediction accuracies were moderate for yield and canning appearance, and high for seed weight, texture, and color retention when individuals from the same breeding cycle were used, and no significant differences were observed between single-trait and multi-trait models, or with the addition of significant SNP markers identified by GWAS in the models. However, when predictions were evaluated between breeding cycles, multi-trait models outperformed single-trait models by up to 61% and 37% for canning appearance and seed yield, respectively, and the use of associated markers identified by GWAS increased prediction accuracy up to 39% and 9% for seed weight and color retention. As genotypes from the new breeding cycle were included in the models, prediction accuracy tended to increase. Our results demonstrate the potential of multi-trait models and GWAS-Assisted genomic prediction to increase the prediction ability of complex traits such as seed yield and canning quality traits in dry beans and exhibit the need of updating the training data set for the implementation of genomic prediction in a dry bean breeding program.

Introduction

Dry bean is a nutrient dense crop and regular consumption has been related with the prevention of cardiovascular diseases in humans (Bogweh and Ageyo, 2021). To increase dry bean consumption, breeders have worked to improve consumer acceptance traits such as cooking time, nutritional and canning quality (Qureshi and Sadohara, 2019). The measure of these traits usually involve specialized methodologies that is often challenging for breeding programs to implement (Mendoza et al., 2014). Canning quality is a measure of how well beans withstand the canning

process and includes attributes such as visual appearance, color, and texture. Canning quality traits are relevant in countries as US, where canned beans represent a major segment of the market of this crop (Miklas, Kelly & Cichy 2022). Despite the importance of canning quality, few studies have been investigated the genetic architecture of the color retention, texture, and appearance of canned dry beans (Wright and Kelly, 2011; Cichy et al., 2014; Sandhu et al., 2018; Bornowski et al., 2020; Bassett et al., 2021), showing the quantitative nature of these traits.

High-throughput phenotyping (HTP) methodologies have proved to be useful to assist selection for complex traits in plant breeding (Rebetzke et al., 2019). In dry beans, the use of near-infrared spectroscopy (NIRS) has been assessed to predict texture, color and visual appearance of canned dry beans, presenting low to moderate accuracy for texture and visual appearance, and high accuracy for color retention (Mendoza et al., 2014; Mendoza et al., 2018). However, predictions of NIRS have been reported with higher accuracy for dietary fiber, uronic acids, calcium, and magnesium when ground dry beans were used compared to intact seeds to take the spectral data (Plans et al., 2012). Nevertheless, from the breeding perspective, the use of non-destructive methodologies such as NIRS are beneficial when there are not available a large number of seeds and/or the number of genotypes is unmanageable with time consuming methodologies. Nowadays, the use of spectral data to predict phenotypes is a rapid and affordable methodology for select complex traits in plant breeding, and has been successfully apply in crops such as soybean, wheat, and maize to predict seed yield and chemical traits (Ge et al., 2019; Hassan et al., 2019; Parmley et al., 2019). However, due to the high dimensionality of spectral data, the use of methods to estimate feature importance for predictors (i.e., wavebands) and variable selection can prevent overfitting and increase the accuracy of prediction (Parmley et al., 2019; Lopez-Cruz et al., 2020).

In addition to HTP, high-throughput genotyping platforms such as genotyping by sequencing (GBS) has been also adopted in plant breeding. GBS is a fast and low-cost methodology to generate hundreds to millions of SNP that have been use in genetic diversity, QTL, genome-wide association studies (GWAS) and genomic prediction in dry beans (Berry et al., 2020; Diaz et al., 2020; Keller et al., 2020; Sadohara et al., 2022). Due the low-cost, GBS is an ideal methodology to use as routine when a breeding program want to implement genomic prediction into the breeding pipeline. Unlike QTL and GWAS methodologies that identify genomic regions that are associated with target traits, genomic prediction uses the information of all the molecular markers in a set of individuals with phenotype and genotype information (training set) to estimate the genetic value

for the target traits of individuals with genotype data (testing set). The relationship between the training and testing sets are of paramount importance for the success of genomic prediction (Cossa et al., 2021). Genomic prediction is useful in quantitative traits, where prediction tools are more suitable than identifying and stack tens to hundreds of QTL in the germplasm (Bernardo, 2020). In sum, genomic prediction has high potential to improve traits that have low heritability, are difficult to measure, or impossible to measure (e.g. not enough seeds, money or equipment to measure the target trait).

All of the genomic prediction studies conducted to date in dry bean have been use single trait (ST) models to estimate the genetic values (Barili et al., 2018; Keller et al., 2020; Diaz et al., 2021a; Diaz et al., 2021b; Shao et al., 2022). However, when phenotypic correlated traits are used as secondary trait using multi-trait (MT) models, the prediction accuracy of MT models can surpass the performance of ST models. (Arojju et al., 2020; Montesinos-Lopez et al., 2022). In addition to correlated traits, selection indices has been also used as secondary traits in MT models and has increased the prediction ability in wheat (Rutkoski et al., 2016).

The integration of genomics and phenomics has the potential to increase the genetic gain per unit of time in plant breeding (Cossa et al., 2021). The objectives of this study were to *i*) compare the performance of ST and MT models in agronomic and canning quality traits in a black dry bean breeding panel, *ii*) assess the used of NIRS to create selection indices to use as secondary trait in MT models, *iii*) evaluate the use of consistent QTL identified through GWAS in genomic prediction models, and *iv*) determine the percentage of new genotypes needed to update genomic prediction after one breeding cycle.

Materials and methods

Plant material

The plant material in this study comprised a population of 415 black breeding lines (BBL) from Michigan State University and USDA-ARS breeding programs. The BBL were evaluated in the experimental field of Michigan State University's (MSU) Saginaw Valley Research and Extension Center in Richville, Michigan. In total, a set of 272 BBL were planted in both 2018 and 2019, while 143 new lines were planted only in 2019. Each genotype was planted in a randomized complete block design with two field replications as 4-row plots with 50 cm row spacing and were end-trimmed to a length of 4.5 m. The black bean cultivars, Eclipse, Zorro, and Zenith were included as checks in both field replications. At harvest maturity, the center two rows of each plot

were harvested, and standard agronomic practices were followed throughout the growing season. The BBL that were planted in both seasons were used as the training set, while the new lines evaluated in 2019 were used as the prediction set to simulate the implementation of GP in a breeding program. The black bean cultivars, Eclipse, Zorro, and Zenith were included as checks in both years. The local standard agricultural practices were followed throughout the growing seasons.

Phenotypic data

The number of days to flowering (DF) was measured as the number of days from planting to when 50% of plants have at least one flower. Days to physiological maturity (DPM) was measured as the number of days from planting to when the first pod begins to discolor in 50% of the plants. Seed weight (SW) was obtained from weighing 100 seeds. Yield (kg ha^{-1}) (YD) was calculated based on the plot size and corrected to a moisture content in the seed of 18%. The dry bean samples were canned and evaluated for appearance, texture and color retention following the protocol described by Wang et al. (2022). Briefly, seed samples were hand-cleaned to remove off-types and split seeds, seed moisture was measured, and 115 g of target solid weight was used. The samples were blanched for 90 seconds at 93.3 °C in tap water with 100 ppm calcium added. After blanching, beans were transferred into cans that were then filled with brine at 93.3 °C. The brine solution contained 15g/L sucrose, 12g/L sodium chloride, and 100 ppm calcium. The cans were cooked in retort (Versatort, Allpax) for 19 minutes at 121.1°C. After two weeks of equilibration at room temperature, the cans were opened for evaluation. Rating of canned beans was conducted by a group of trained panelists with a scale of 1-5 for appearance (App) (1: severe split seeds, 5: <10% of split seeds) and color (Col) (1: lightest, 5: darkest). In addition, a Hunter Labscan XE colorimeter was used to measure CIE (International Commission on Illumination) L*, a*, and b* values for each sample. L* measures darkness to lightness from 0 (black) to 100 (white). a* measures the level of greenness (negative values) to redness (positive values). b* measures the level of blueness (negative values) to yellowness (positive values). Furthermore, the texture (Text) of processed samples was measured as the peak force required to completely compress a 100 g of washed and drained beans using a texture analyzer with a Kramer Shear press attachment (Texture Technologies Corp., USA). The rows and columns from the field were used as random effects to fit a linear mixed model using the function “SpATS” and “SAP” for the R package SpATS

(Rodríguez-Álvarez et al., 2018). The genotype was fitted as fixed effect to obtain the best linear unbiased estimators (BLUE) for each trait.

Genotypic data

DNA was extracted from trifoliolate leaves using NucleoSpin Plant II Kit (Macherey–Nagel, Duren, Germany) following the ‘Genomic DNA from plant’ protocol. DNA concentration was measured by using Quant-iT™ PicoGreen™ dsDNA Assay Kit (Invitrogen, Waltham, MA, USA), and 10 ng/μL of DNA was used for GBS library preparation according to Elshire et al. (2011) with a single restriction enzyme, ApeKI. Each plate of 96-wells was sequenced in a lane of an Illumina HiSeq platform using single-end reads at RTSF Genomics Core at Michigan State University.

The libraries were demultiplexed using NGSEP (v3.1.2) (Tello et al., 2019). Adapters and low-quality bases from the raw sequencing data were trimmed using Cutadapt v 1.16, and the processed reads were aligned to the reference genome of *P. vulgaris* v2.1 G19833 (Schmutz et al., 2014) using Bowtie2 (v2.2.30) (Langmead and Salzberg, 2012) with default parameters. The SNP calling was carried out by using NGSEP software following recommended parameters for GBS data (Perea et al., 2016). The merged genotypic matrix was filtered with NGSEP for variants that were in the predicted repetitive regions of the reference genome *P. vulgaris* v2.1 by Lobaton et al. (2018), non-biallelic, genotype quality above 30, a maximum observed heterozygosity of 0.05 per SNP, more than 20% of missing data per site, and minor allele frequency (MAF) above 0.01. Besides, SNPs in linkage disequilibrium above 0.9 using a window of 500 SNPs were removed using Bcftools (Li, 2011). The resulting genotype matrix was imputed using Beagle V5.4 (Browning et al., 2018).

Near-infrared spectroscopy (NIRS) and Regularized selection indices (RSI)

Intact dry beans were scanned using a near-infrared reflectance diode array feed analyzer (Pertec, Springfield, IL, USA) in absorbance mode in the range of 1100 to 2200 nm collected at increments of 2 nm. For scanning, 80 g of dry beans were placed on a Petri dish and the spectrum of each sample was the average of 50 scans. The spectra was preprocessed by Standard Normal Variate (SNV) using the R package prospectr (Stevens and Ramirez–Lopez, 2022). The preprocessed spectral data was used to build the regularized selection indices (RSI) with the following equation: $RSI_i = x_i' \beta$, where $x_i = (x_{i1}, \dots, x_{ip})$, is a vector of wavelengths and $\beta = (\beta_1, \dots, \beta_p)'$ is a vector of regression coefficients of each wavelength. Using a Ridge-regression-

type penalized selection indices proposed by (Lopez-Cruz et al., 2020), regulation is achieved by including the penalty parameter λ to avoid overfitting caused for high-dimensional data such as NIRS through variable selection (i.e., RGS based in a subset of the wavelengths):

$$\hat{\beta} = (P_x + \lambda I)^{-1} G_{x,y}$$

Where P_x is the ‘phenotypic’ variance-covariance matrix of the wavelengths, I is an identity matrix, and G is a vector of the genetic covariances between the target trait (y_i) and each wavelength (x_i). After the estimation of $\hat{\beta}$, the RSI is a sum of the selected wavelengths with their regression coefficients estimated to maximize the correlation between the target trait and the RSI. To assess the prediction accuracy of RSI, the training set was divided into training and validation subsets by randomly assigning 70% and 30% of the individuals, respectively. To optimize the penalization value λ , a 10-fold cross validation was carried out in the training subset, and RSI was derived over a grid of 100 values of λ . The accuracy measured as the Pearson correlation between RSI and each trait was used to identify the value of λ that maximized accuracy in each cross validation. The optimal value of λ was defined as the average value of λ across each cross-validation and was used to predict the validation subset. The training-validation procedure described above was repeated 100 times by randomly assigning samples from the training set into training and validation subsets, and was implemented using the R package SFSI (Lopez-Cruz et al., 2020).

Genome-wide association study (GWAS)

GWAS was conducted using the multi-locus methods Bayesian-information and linkage-disequilibrium iteratively nested keyway (BLINK) and Fixed and random model Circulating Probability Unification (FarmCPU) implemented in the R package GAPIT v3 (Wang and Zhang, 2021). GWAS was conducted with all data collected in 2018 and 2019, and a Bonferroni threshold of $\alpha=0.05$ was used to identify associations. The QTL associations were defined as consistent when the same SNP was identified in both years, or when two different SNPs were associated in the same genomic region between years, and the distance between them was not greater than 2 Mb.

Genomic prediction

Genomic prediction (GP) was assessed using the same 100 partitions described above in the RSI. Single trait (ST) and multi-trait (MT) models were performed using the Bayesian

Reproducing Kernel Hilbert Spaces (RKHS) with 10,000 iterations and 1,000 burn-in using the R package BGLR (Pérez-Rodríguez and de Los Campos, 2022). The MT models were implemented assuming the marker-based relationship as an unstructured matrix and the residual variance-covariance between traits as diagonal matrix. In addition to the target trait to predict, MT models included the RSI, and for seed yield and appearance, seed weight and texture were used, respectively, due to their positive correlation with the target trait across years. Furthermore, the effect of consistent QTL identified in GWAS were used as fixed variables in the models. Prediction ability is expressed as a Pearson correlation coefficient between the observed and predicted values in the validation subset.

Prediction in the testing population

To simulate the implementation of GP in the dry bean program, the new BBL included in 2019 were used as a prediction set. The prediction accuracy of five models were assessed using the phenotypic data from the training (272 genotypes) and prediction (143 genotypes) sets from 2019: *i*) ST, *ii*) ST using QTL as fixed variables, *iii*) MT using the RSI for the target trait, *iv*) MT using correlated traits, and *v*) MT using RSI, correlated traits, and QTL as fixed variables. Additionally, as the genetic relationship change between breeding cycles, 10% (14), 20% (29), 30% (43), 40% (57) and 50% (72) of the genotypes in the prediction set were randomly assigning to the training population to assess the number of genotypes needed to update the GP models. The procedure described above was repeated 100 times, and was implemented using the R package BGLR (Pérez-Rodríguez and de Los Campos, 2022)

Results

Phenotypic data

The black bean breeding lines were grown over two years in Michigan with 272 lines grown in both 2018 and 2019 and an additional set of 143 lines added in 2019. The agronomic traits including DF, DPM, SW, and YD showed strong year to year variation (**Supplementary Figure 3.1**). The variation across years was attributed to higher rainfall during the growing season in 2018 compared to 2019. The canning quality traits including appearance, color retention, and $L^*a^*b^*$ color values were more consistent across years, with the exception of texture which was lower in 2018 compared to 2019 (**Supplementary Figure 3.2**). A positive correlation was observed between YD and the other agronomic traits (DF, DPM, SW) in both growing seasons, but in 2018 when the precipitation was higher, the correlations were stronger, especially the correlation

between YD and DPM which was 0.43 ($p < 0.001$) and 0.08 ($p > 0.05$) in 2018 and 2019, respectively (**Supplementary Figure 3.3**). Canned bean appearance ratings made by panelists were positively correlated with texture (2018: $r = 0.30$, 2019: $r=0.20$ with $p < 0.001$) and color (2018: $r = 0.38$, 2019: $r=0.24$ with $p < 0.001$). Furthermore, L^* a^* and b^* color values were highly correlated with color ratings made by panelists in both years, with r values ranging from -0.73 to -0.82 within the same year and from -0.58 to -0.66 when comparing across years.

Genotypic data

In total, 2,315 SNP markers were retained after filtering based on missing data, minor allele frequency, and LD. The genomic relationship matrix showed a different degree of genetic relationship between families representing the use of common parents in the breeding populations, and the genomic relationship difference was higher in the training set showing two subgroups representing the two breeding programs (MSU vs ARS) (**Figure 3.1 A**). The genetic diversity was evaluated through principal component analysis (PCA). Generally, there was low structure within the population with varying degrees of admixture which is expected from a panel of breeding lines derived from different bi-parental populations and breeding cycles (**Figure 3.1 B**).

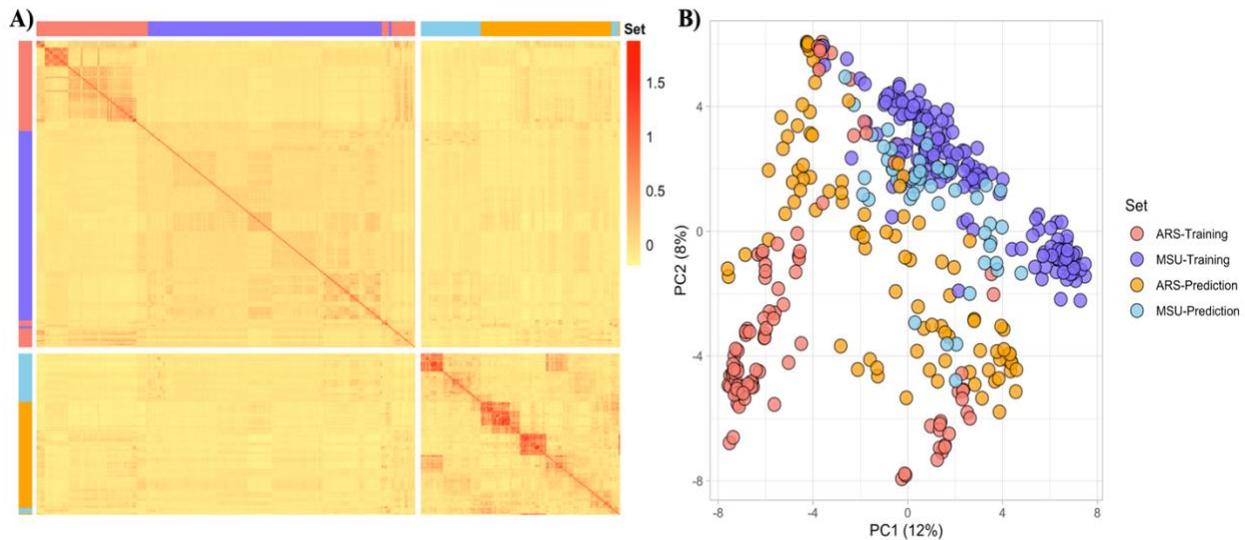


Figure 3.1 Genetic structure of the black bean breeding panel. A Heatmap of the genomic relationship matrix. B First two principal components showing the location of each genotype defined by the eigenvectors. The colors in the PCA and axes of the heatmap represent the genotypes in the training and prediction sets from the MSU and ARS breeding programs.

Genome-wide association study (GWAS)

Marker-trait associations were evaluated using the multi-marker approaches FarmCPU and BLINK. GWAS peaks were selected for those SNPs with a significance greater than that established by the Bonferroni correction (2.16×10^{-5}) and only QTL consistently associated across years are reported (**Supplementary table 3.1**). In total, 20 associations were found in both years, 13 for canning quality and 7 for agronomic traits. A robust QTL was found in chromosome 2 at 48.4 Mb related to all three L*, a*, b* color components (**Supplementary Figure 3.5**), however, it was not associated with the panelist color ratings. Two regions on chromosome 8 (7.2 Mb) and 11 (53.1 Mb) were associated with color ratings (**Figure 3.2**). Furthermore, four associations were found for texture in chromosomes 2 (38.3 Mb), 5 (39.4-39.6 Mb), 7 (9.4 Mb) and 10 (43.2-43.6 Mb) (**Figure 3.2**). For appearance, four associations were identified in chromosomes 2 (46.7 Mb), 3 (52.6-53.0 Mb), 5 (0.6-1.3 Mb), and 9 (14.0-14.5 Mb) (**Figure 3.2**). Additionally, seven associations for agronomic traits were found across years (**Supplementary table 3.1, Supplementary Figure 3.6**).

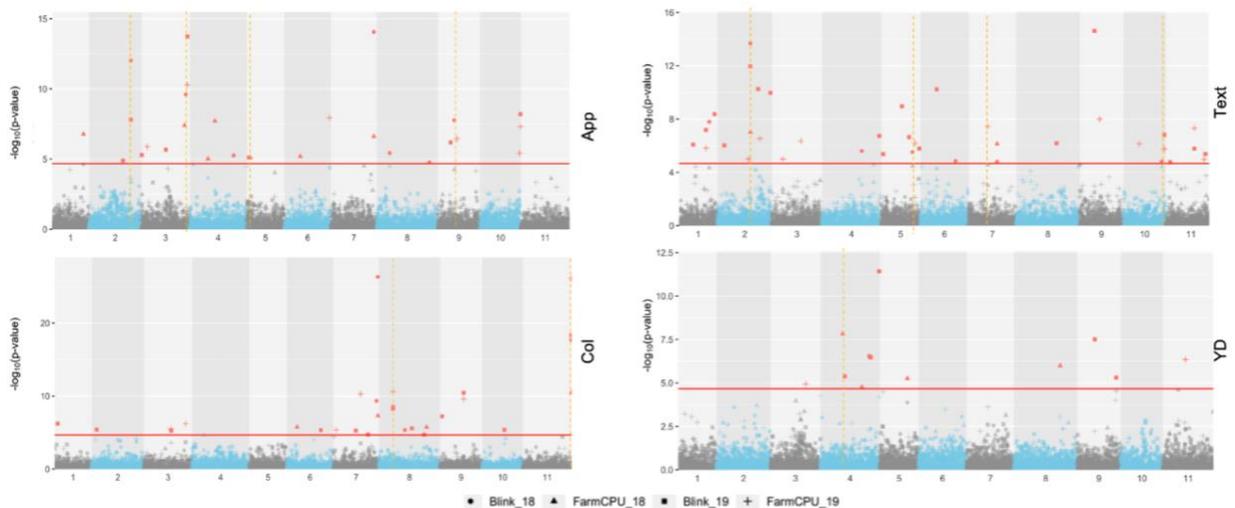


Figure 3.2 Genome-wide association analysis of seed yield, color (Col), appearance (App), and texture (Text). The red lines indicate the Bonferroni threshold ($\alpha = 0.05$), and the vertical yellow dotted lines indicate common QTL for 2018 and 2019.

Regularized selection indices and Genomic prediction - training population

In total, 100 partitions were used to assess the prediction accuracy for the regularized selection indices (RSI) of the NIRS data and Genomic Prediction (GP) models in the training set composed of 272 BBL. The prediction accuracies of RSI, single, and multi-trait GP models are presented in **Figure 3.3** and **Supplementary Figure 3.7**. RSI exhibited the lowest prediction accuracies for all

traits, while genomic prediction models showed variable prediction accuracies ranging from 0.32 for DPM in 2019 (when QTL were used as fixed variables) to 0.92 for texture in 2019. Overall, genomic prediction accuracies were consistent across years, except for texture which in 2018 had an accuracy for single and multi-trait models of 0.57, while in 2019 the accuracy increased up to 0.92. The accuracy for texture was related to the higher heritability observed in 2019 ($h^2:0.94$) compared to 2018 ($h^2:0.80$) **Supplementary Table 3.2**. Prediction accuracies were similar between ST and MT models, although there was a small improvement in YD and App when the correlated traits SW and texture, respectively were used in the model. Moreover, the positive effect of QTL was not consistent across traits, except for color. ST model using fixed QTL improved prediction accuracies for color in 2018 (0.03) and 2019 (0.02) compared to ST models without QTL information.

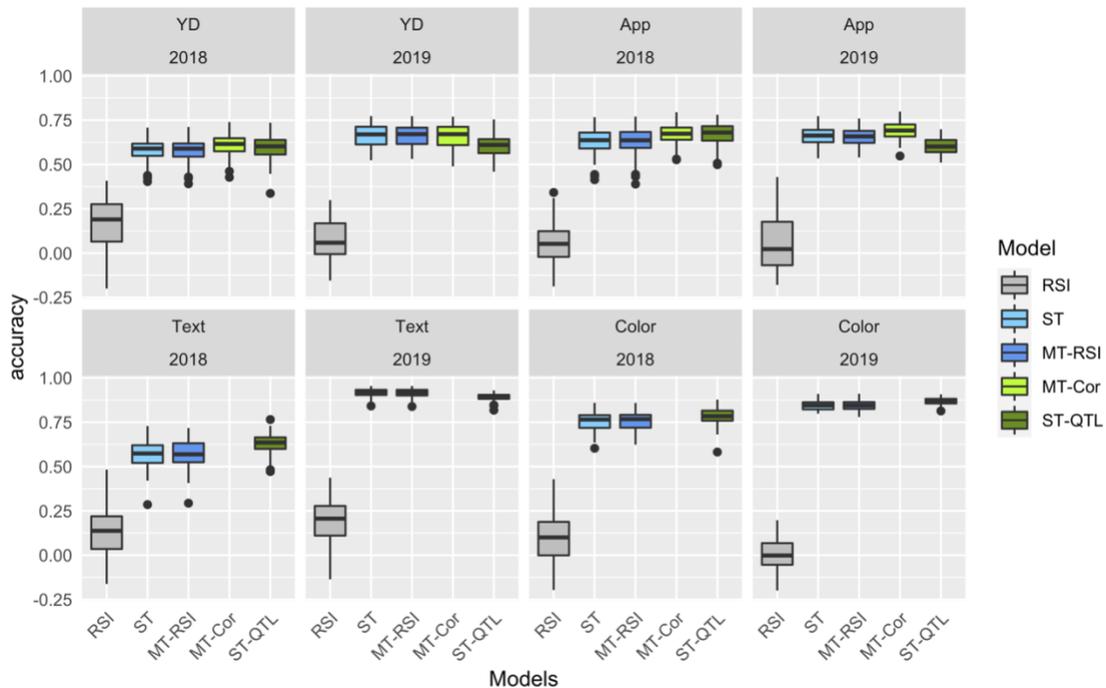


Figure 3.3 Prediction accuracy of seed yield and canning quality traits in the BBL using RSI, single-trait (ST) and multi-trait (MT) models. The distribution of boxplots represents 100 partitions in the training set.

Regularized selection indices and Genomic prediction - testing population

To simulate the implementation of GP in the breeding program, 143 genotypes belonging to a preliminary yield trial grown in 2019 were used as a testing set. The prediction accuracy of ST and MT models adding different proportions (from 0 to 50%) of the testing set to train the models are

presented in **Figure 3.4-5** and **Supplementary Figure 3.8-9**. Overall, the lowest prediction accuracies were observed when no data points from the testing set were included to train the model, and the accuracy tended to increase with the addition of data points belonging to the same breeding cycle. When 30% of genotypes were added from the testing set in ST the prediction accuracy increased by 61% (0.22 vs 0.36) in YD and by 12% in App (0.31 vs 0.37). Furthermore, MT boosted the prediction ability in YD and App using SW and texture in the multivariate model (MT-Cor) **Supplementary Table 3.3**. Adding 30% of individuals from the testing set and correlated traits into the model increased 82% (0.22 vs 0.40), and 52% (0.31 vs 0.47) the prediction accuracy of YD and App, respectively. However, the highest prediction accuracy for App was achieved with MT-Cor without adding new genotypes (0.50). For YD and App, the MT-Cor model yielded the best prediction accuracy without regard to the percentage of the testing population used to train the model.

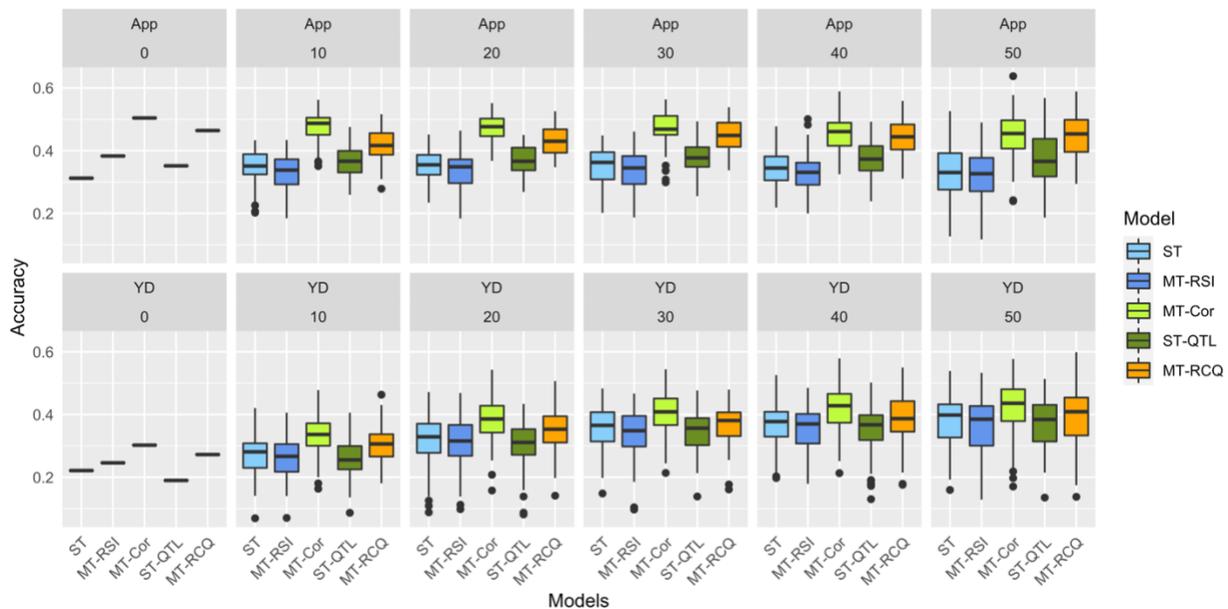


Figure 3.4 Prediction accuracy of appearance (App) and Yield (YD) in the 2019 testing set comprised 143 black breeding lines. Different proportions of the testing set were included in the models (0%, 10% =14, 20% = 29, 30%=43, 40%= 57, 50%=72). The distribution of boxplots represents 100 partitions.

In general, the use of QTL as fixed effects did not show a strong effect on prediction accuracies in most of the traits. However, the use of QTL increased the predictions of Col and SW (**Figure 3.5, Supplementary Figure 3.8**), while reducing the prediction ability for b* (**Supplementary**

Figure 9). The use of RSI in multi-trait models (MT-RSI) only increased the prediction accuracy of texture, and the prediction was boosted when MT-RSI used QTL as fixed effects (**Figure 3.5**).

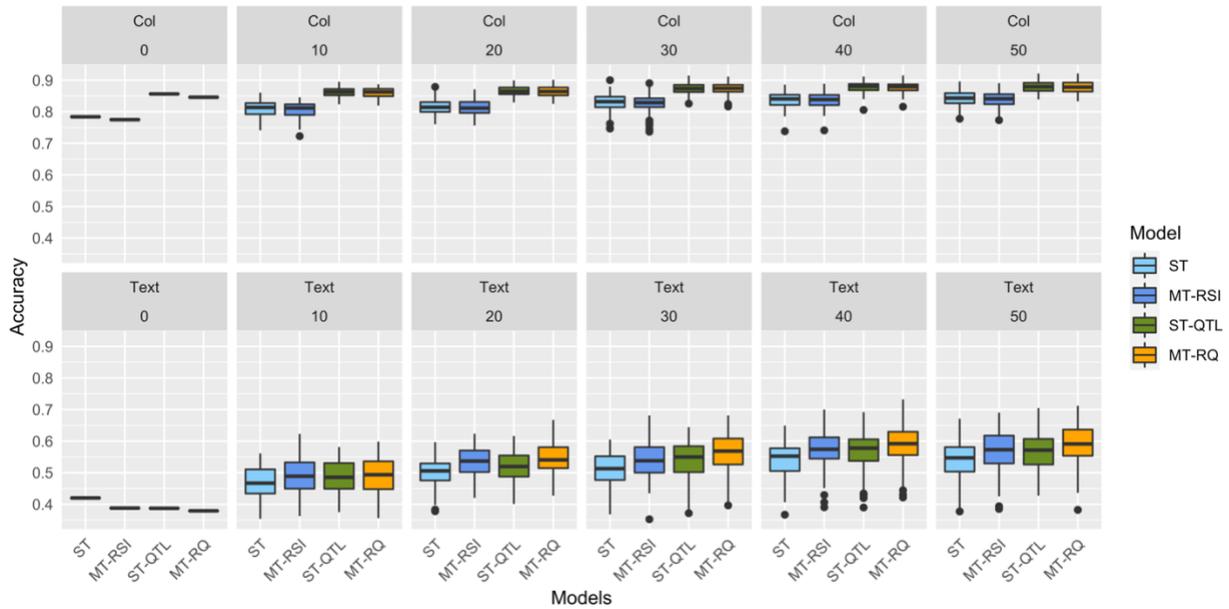


Figure 3.5 Prediction accuracy of color (Col) and texture (Text) in the 2019 testing set comprised of 143 black breeding lines. Different proportions of the testing set were included in the models (0%, 10% =14, 20% = 29, 30%=43, 40%= 57, 50%=72). The distribution of boxplots represents 100 partitions.

Discussion

In the current study, the genetic architecture of agronomic and canning quality traits was dissected and used in genomic prediction models to assess their potential use in dry bean breeding. In general, prediction accuracies were high in genotypes from the same breeding cycle (training-validation partitions). This is expected due that RILs from the same breeding cycle tend to have more parents in common. In overall, traits with high h^2 also had higher prediction accuracies (e.g., texture, color and seed weight), while traits with lower heritability, including seed yield and canning appearance presented lower prediction accuracies. Days to flowering and days to maturity are oligogenic traits with low genetic complexity. However, the measurement of these traits is subject to error based on measurement timing. This can increase the noise of data, reducing the phenotypic variance explained by genetics (h^2) and at the end, reducing the prediction ability of genomic tools. This can be observed in days to maturity, where the h^2 changed from 0.83 in 2018 to 0.26 in 2019. Seed yield, DF and DM were the traits with lower h^2 compared with the other traits and showed moderate prediction abilities in the training-validation partitions. Although

appearance and color are measured by trained panelists with a visual scale, and they are subjective as are DF and DM, the conditions in the lab are more uniform than in the field and the score per sample is an average of the readings of each trained panelist (8-12 people), which helps to reduce the subjectivity of these ratings. The h^2 of both traits (appearance and color) was high in the two growing seasons and prediction abilities for color rating were similar to L^* a^* and b^* using ST models. In the training-validation partitions the average of prediction accuracies for the traits ranged from 0.44 in days to flowering to 0.92 in texture using the ST models, which could be considered as moderate to high prediction accuracy. This study confirms the promising results for genomic prediction found by Keller et al. (2020) in an Andean elite bean breeding for traits with complex genetic architecture. Using twelve field trials, Keller et al. (2020) reported a prediction accuracies up to 0.6 for yield, and a similar prediction accuracy was observed in this study for yield using the ST model in 2018 (0.58) and 2019 (0.66).

Updating training population

The genetic relationship between training and the testing sets is of paramount importance for increasing the accuracy of genomic prediction (Bernardo, 2020). It is well-reported, that updating and increasing the training population tends to improve prediction accuracy (Spindel et al., 2016; Lopez-Cruz et al., 2021). In the current study, 143 genotypes belonging to a preliminary yield trial in 2019 were used as the testing set and the training set was updated with 10 to 50% of the new genotypes to train the models. Prediction accuracies consistently increase with the percentage of new genotypes included to train the ST model for most of the traits, increasing up to 2.6-fold prediction accuracies for days to maturity and 72% of the predictions for seed yield when 50% of the testing set was included in the model. Most of the prediction gain was reached when 30% of the testing set was used in the model, which confirms the well-established evidence that updating the training set increase the genetic relationship between the training and testing set, which produces better predictions. However, the improvement was different across traits, which suggests that improvement in prediction is also linked to the genetic architecture of each trait. The a^* was the trait with the lowest improvement of prediction when the ST used genotypes form the testing set (0.004%). However, the prediction accuracy without adding new genotypes in the training population was 0.74 for this trait, which suggest that, although absence of strong genetic relationship between training and testing sets, this is an easy trait to predict using ST models.

Multi trait models

Genomic selection models exploit the genetic relationship between genotypes to predict the genetic values, the higher the genetic relationship between the training and testing sets, the better the prediction ability. In addition to the genetic relationship among genotypes, the MT models use the information of phenotypic correlated traits, and similar to the relationship between genotypes, the larger the correlation between traits in the model, the better the prediction ability of MT models (Montesinos-Lopez et al., 2022). In this study, the use of MT models resulted in a similar prediction ability as the ST models when applied within the training set comprised of 272 black genotypes from the same breeding cycle. Lines from the same breeding cycle tend to share more parents and have a higher genetic relationship compared with other breeding cycles. However, MT increased the prediction ability to estimate the genetic value in the testing population from a different breeding cycle. Using correlated phenotypes as secondary traits in a multivariate model increased the prediction ability of appearance and seed yield. As reported in previous studies the extent of improvement of predictions depends on the correlation between traits and the heritability of the secondary trait (Arojju et al., 2020).

The secondary traits (seed weight and texture) showed a high h^2 with a moderate correlation with target traits, and in both cases, they increased prediction accuracy in the training and testing sets. Seed weight is an important yield component, and although dry bean breeders are locked in to a standard seed size for each market class, seed weight is a fast and easy phenotype to collect, has high heritability, and tends to have a positive correlation with seed yield. For canning traits, texture increased more than 60% the prediction ability for appearance in the testing population. The prediction ability of ST model for appearance did not increase significantly when genotypes from the new breeding cycle were included in the training set, and surprisingly, the highest prediction ability in the testing set for appearance was achieved when the texture was used as a secondary trait without adding new genotypes to train the model. These results show the importance of texture as a secondary trait to predict appearance, however, texture is a laborious trait that was taken in canned beans.

Non-destructive high-throughput phenotyping methodologies such as NIRS have been reported to be used successfully in plant breeding (Jiang, 2020; Masilamani et al., 2020). However, some studies have reported moderate to low prediction accuracy for appearance and texture in canned dry beans using NIRs (Mendoza et al., 2014), and also it is reported that the prediction of

seed components has a lower accuracy in intact seed compared to ground seed in beans (Plans et al., 2012). In the current study, the use of NIRS to estimate a regularized selection indices using intact seeds has a low correlation with target traits and did not improve prediction accuracies when included as a secondary trait in MT models in the training-validation partitions. However, the use of RSI in multi-trait models (MT-RSI) increased the prediction accuracy of texture in the testing set, and the prediction ability was boosted when MT-RSI used QTL as fixed effects. Using several spectral preprocessing methods, Mendoza et al. (2018) reported a significant improvement of predictions of texture using NIRS. In the current study, the standard normal variate (SNV) was used to perform a normalization of the spectral data, and due to the low correlation between indices and traits, other spectral preprocessing methods such as derivatives and multi-resolution wavelet could be more beneficial to increase the accuracy of selection indices using NIRS. Although a tuning process is needed to preprocess and build selection indices using high dimensional data such as NIRS, the use of high-throughput phenotyping has potential to generate secondary traits to increase the prediction ability of canning quality traits using MT models.

GWAS-Assisted genomic prediction

The use of markers linked to QTL as fixed effects improve the estimation of their effects, while the remaining markers are adjusted for the contribution of fixed QTL (Bernardo, 2014; Bernardo, 2020). In the current analysis, several genomic regions were associated with target traits in both years and a small number of consistent QTL across growing seasons were found (**Supplementary Table 3.1**). Two association for texture in chromosome 7 (9,4 Mb) and 10 (43.2-43.6 Mb) were reported in previous studies (Cichy et al., 2014; Bassett et al., 2021). The association found on chromosome 7 overlaps with the Asp gene that has been related to water uptake (Cichy et al., 2014). Four associations were identified for appearance, and one in chromosome 5 (0.6-1.3 Mb) was proximal to a QTL reported by Wright & Kelly (2011). For color, two QTL were identified in chromosome 8 (7.2 Mb) and chromosome 11 (53.1 Mb), these associations were proximal to QTL for L* and b* reported in a QTL analysis in black beans (Bornowski et al., 2020). L* measures darkness to lightness, while b* measures the level of blueness to yellowness, and the correlation of L* and b* with color was < -0.78 , which shows the relationship between these traits. From the seven associations identified for agronomic traits, six were identified in the meta-analysis of QTL and GWAS of seed yield components reported by Izquierdo et al., (under review). The consistent association for yield in chromosome 4 (1.6 – 1.8 Mb) overlapped with MQTL-YC4.1 (Izquierdo

et al., under review). MQTL-YC4.1 was supported by QTL related to seed yield, seed weight, seeds per pod, and pod harvest index, which highlights the importance of this region in dry bean breeding. The three QTL related with SW co-localized with MQTL3.3 (supported by SW and number of pods per plant), MQTL4.5 (supported by yield and number of seeds per plant), and MQTL6.2 (supported by SW, pod harvest index, and yield) reported by Izquierdo et al. (under review).

The use of market-assisted selection using QTL information for quantitative traits could be challenging due to the high number of markers needed, their small effect, and their interaction with the environment. An alternative use of QTL is to use them as fixed effects in genomic prediction models. We identified that in most cases the use of consistent QTL in genomic prediction models does not reduce prediction abilities and in cases such as color and seed weight the prediction increases as was observed in the training-validation partitions and in the testing set. However, fixed QTL reduces prediction ability in DM and b^* . In DM this reduction could be explained by the low h^2 in 2019, as reported by Bernardo (2014), QTL can increase prediction ability only when h^2 range from moderate to high. Besides, the reduction of prediction ability using QTL for b^* could be the result of using a false association. Although L^* , a^* and b^* all showed consistent associations at the end of chromosome 2 (48.4 Mb) in 2018, this association was present in a different position in 2019 (**Supplementary Table 3.1**). In 2018 using the training set, the SNP peak associated with b^* located at 48.4 Mb increased prediction ability, while in 2019 the SNP peak at 47.6 Mb reduced prediction ability in the training and prediction set, which suggests that the association identified in 2019 is near but is not the QTL related to b^* . In sum, the use of QTL as fixed effects increased the prediction accuracy of texture, color, seed weight, and a^* , while reducing the prediction accuracy of DM and b^* . In the other traits the prediction ability was not affected with the inclusion of fixed markers.

Conclusions

Using an affordable genotyping methodology such as GBS, we were able to get moderate (0.44) to high predictions (0.92) for agronomic and canning quality traits in a black bean breeding panel from the same breeding cycle, when the genetic relationship tends to be high due to the use of common parents. When individuals from another breeding cycle were used as prediction set, the accuracy of genomic prediction dropped, and factors such as the inclusion of phenotypic correlated traits in multi-trait models, QTL as fixed effects, and updating the training set with

individuals from the new cycle are useful to improve the estimation of the breeding values, yielding in higher prediction accuracies. Canning quality traits are laborious and costly traits that are measured in advanced generations. We observed that with enough genetic relationship between training and testing populations such as the observed in the training set from this study, the prediction accuracies for canning quality traits are higher than 0.57, which could be high enough to implement genomic prediction in early generations to increase selection intensity, leading to higher genetic gain for these traits.

BIBLIOGRAPHY

- Arojju SK, Cao M, Trolove M, Barrett BA, Inch C, Eady C, Stewart A, Faville MJ (2020a) Multi-Trait Genomic Prediction Improves Predictive Ability for Dry Matter Yield and Water-Soluble Carbohydrates in Perennial Ryegrass. *Front Plant Sci* 11: 1
- Arojju SK, Cao M, Trolove M, Barrett BA, Inch C, Eady C, Stewart A, Faville MJ (2020b) Multi-Trait Genomic Prediction Improves Predictive Ability for Dry Matter Yield and Water-Soluble Carbohydrates in Perennial Ryegrass. *Front Plant Sci* 11: 1–19
- Barili LD, Vale NM do, Silva FF e, Carneiro JE de S, Oliveira HR de, Vianello RP, Valdisser PAMR, Nascimento M (2018) Genome prediction accuracy of common bean via Bayesian models. *Ciência Rural* 48: 1–6
- Bassett A, Katuramu DN, Song Q, Cichy K (2021) QTL Mapping of Seed Quality Traits Including Cooking Time, Flavor, and Texture in a Yellow Dry Bean (*Phaseolus vulgaris* L.) Population. *Front Plant Sci* 12: 1–16
- Bernardo R (2014) Genomewide Selection when Major Genes Are Known. *Crop Sci* 54: 68–75
- Bernardo R (2020) *Breeding for Quantitative Traits in Plants*. Stemma Press. doi: 10.2135/cropsci2003.1578
- Berry M, Izquierdo P, Jeffery H, Shaw S, Nchimbi-Msolla S, Cichy K (2020) QTL analysis of cooking time and quality traits in dry bean (*Phaseolus vulgaris* L.). *Theor Appl Genet*. doi: 10.1007/s00122-020-03598-w
- Bogweh NE, Ageyo OC (2021) Do common beans (*Phaseolus vulgaris* L.) promote good health in humans a systematic review and meta-analysis of clinical and randomized controlled trials. *Nutrients*. doi: 10.3390/nu13113701
- Bornowski N, Song Q, Kelly JD (2020) QTL mapping of post-processing color retention in two black bean populations. *Theor Appl Genet* 1: 3
- Browning BL, Zhou Y, Browning SR (2018) A One-Penny Imputed Genome from Next-Generation Reference Panels. *Am J Hum Genet* 103: 338–348
- Cichy KA, Fernandez A, Kilian A, Kelly JD, Galeano CH, Shaw S, Brick M, Hodkinson D, Troxtell E (2014) QTL analysis of canning quality and color retention in black beans (*Phaseolus vulgaris* L.). *Mol Breed* 33: 139–154
- Crossa J, Fritsche-Neto R, Montesinos-Lopez OA, Costa-Neto G, Dreisigacker S, Montesinos-Lopez A, Bentley AR (2021) The Modern Plant Breeding Triangle: Optimizing the Use of Genomics, Phenomics, and Enviromics Data. *Front Plant Sci* 12: 1–6
- Diaz LM, Arredondo V, Ariza-Suarez D, Aparicio J, Buendia HF, Cajiao C, Mosquera G, Beebe SE, Mukankusi CM, Raatz B, et al (2021a) Genetic Analyses and Genomic Predictions of Root Rot Resistance in Common Bean Across Trials and Populations. *Front Plant Sci* 12: 1–15
- Diaz S, Ariza-Suarez D, Izquierdo P, David Lobaton J, Fernando de la Hoz J, Acevedo F, Duitama J, Guerrero AF, Cajiao C, Mayor V, et al (2020) Genetic mapping for agronomic traits in a

- MAGIC population of common bean (*Phaseolus vulgaris* L.) under drought conditions. doi: 10.1186/s12864-020-07213-6
- Diaz S, Ariza-Suarez D, Ramdeen R, Aparicio J, Arunachalam N, Hernandez C, Diaz H, Ruiz H, Piepho H-P, Raatz B (2021b) Genetic Architecture and Genomic Prediction of Cooking Time in Common Bean (*Phaseolus vulgaris* L.). *Front Plant Sci* 11: 1–16
- Elshire RJ, Glaubitz JC, Sun Q, Poland J a, Kawamoto K, Buckler ES, Mitchell SE (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One* 6: e19379
- Hassan MA, Yang M, Rasheed A, Yang G, Reynolds M, Xia X, Xiao Y, He Z (2019) A rapid monitoring of NDVI across the wheat growth cycle for grain yield prediction using a multi-spectral UAV platform. *Plant Sci* 282: 95–103
- Keller B, Ariza-Suarez D, de la Hoz J, Aparicio JS, Portilla-Benavides AE, Buendia HF, Mayor VM, Studer B, Raatz B (2020) Genomic Prediction of Agronomic Traits in Common Bean (*Phaseolus vulgaris* L.) Under Environmental Stress. *Front Plant Sci* 11: 1001
- Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9: 357–9
- Li H (2011) A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* 27: 2987–2993
- Lobaton JD, Miller T, Gil J, Ariza D, Hoz JF, Soler A, Beebe S, Duitama J, Gepts P, Raatz B (2018) Resequencing of Common Bean Identifies Regions of Inter-Gene Pool Introgression and Provides Comprehensive Resources for Molecular Breeding. *Plant Genome* 11: 170068
- Lopez-Cruz M, Beyene Y, Gowda M, Crossa J, Pérez-Rodríguez P, de los Campos G (2021) Multi-generation genomic prediction of maize yield using parametric and non-parametric sparse selection indices. *Heredity (Edinb)* 127: 423–432
- Lopez-Cruz M, Olson E, Rovere G, Crossa J, Dreisigacker S, Mondal S, Singh R, Campos G de los (2020) Regularized selection indices for breeding value prediction using hyper-spectral image data. *Sci Rep* 10: 1–12
- Mendoza FA, Cichy K, Lu R, Kelly JD (2014) Evaluation of Canning Quality Traits in Black Beans (*Phaseolus vulgaris* L.) by Visible/Near-Infrared Spectroscopy. *Food Bioprocess Technol* 7: 2666–2678
- Mendoza FA, Cichy KA, Sprague C, Goffnett A, Lu R, Kelly JD (2018) Prediction of canned black bean texture (*Phaseolus vulgaris* L.) from intact dry seeds using visible/near infrared spectroscopy and hyperspectral imaging data. *J Sci Food Agric* 98: 283–290
- Montesinos-Lopez OA, Montesinos-Lopez A, Mosqueda-Gonzalez B, Montesinos-Lopez J, Crossa J (2022) Genomic Prediction of Complex Traits. *Methods in molecular biology*
- Perea C, De La Hoz JF, Felipe Cruz D, Lobaton JD, Izquierdo P, Quintero JC, Raatz B, Duitama J (2016) Bioinformatic analysis of genotype by sequencing (GBS) data with NGSEP. *BMC*

- Pérez-Rodríguez P, de Los Campos G (2022) Multitrait Bayesian shrinkage and variable selection models with the BGLR-R package. *Genetics*. doi: 10.1093/genetics/iyac112
- Plans M, Simó J, Casañas F, Sabaté J (2012) Near-infrared spectroscopy analysis of seed coats of common beans (*Phaseolus vulgaris* L.): A potential tool for breeding and quality evaluation. *J Agric Food Chem* 60: 706–712
- Qureshi AMI, Sadohara R (2019) Quality breeding in field crops. *Qual Breed F Crop*. doi: 10.1007/978-3-030-04609-5
- Rodríguez-Álvarez MX, Boer MP, van Eeuwijk FA, Eilers PHC (2018) Correcting for spatial heterogeneity in plant breeding experiments with P-splines. *Spat Stat*. doi: 10.1016/j.spasta.2017.10.003
- Rutkoski J, Poland J, Mondal S, Autrique E, González Pérez L, Crossa J, Reynolds M, Singh R (2016) Canopy Temperature and Vegetation Indices from High-Throughput Phenotyping Improve Accuracy of Pedigree and Genomic Selection for Grain Yield in Wheat. doi: 10.1534/g3.116.032888
- Sadohara R, Izquierdo P, Couto Alves F, Porch TG, Beaver JS, Urrea CA, Cichy KA (2022) The *Phaseolus vulgaris* L. Yellow Bean Collection: Genetic diversity and characterization for cooking time (Under review). *Genet Resour Crop Evol*. doi: 10.1007/s10722-021-01323-0
- Sandhu KS, You FM, Conner RL, Balasubramanian PM, Hou A (2018) Genetic analysis and QTL mapping of the seed hardness trait in a black common bean (*Phaseolus vulgaris*) recombinant inbred line (RIL) population. *Mol Breed*. doi: 10.1007/s11032-018-0789-y
- Schmutz J, McClean PE, Mamidi S, Wu GA, Cannon SB, Grimwood J, Jenkins J, Shu S, Song Q, Chavarro C, et al (2014) A reference genome for common bean and genome-wide analysis of dual domestications. *Nat Genet* 46: 707–713
- Shao J, Hao Y, Wang L, Xie Y, Zhang H, Bai J, Wu J, Fu J (2022) Development of a Model for Genomic Prediction of Multiple Traits in Common Bean Germplasm, Based on Population Structure. *Plants*. doi: 10.3390/plants11101298
- Spindel JE, Begum H, Akdemir D, Collard B, Redoña E, Jannink J-L, McCouch S (2016) Genome-wide prediction models that incorporate de novo GWAS are a powerful new tool for tropical rice improvement. *Heredity (Edinb)* 116: 395–408
- Stevens A, Ramirez-Lopez L (2022) An Introduction to the Psychotherapies. *R Packag Vignette R Packag version 026*. doi: 10.1176/ajp.136.12.1628-a
- Tello D, Gil J, Loaiza CD, Riascos JJ, Cardozo N, Duitama J (2019) NGSEP3: Accurate variant calling across species and sequencing protocols. *Bioinformatics* 35: 4716–4723
- Wang J, Zhang Z (2021) GAPIT Version 3: Boosting Power and Accuracy for Genomic Association and Prediction. *Genomics, Proteomics Bioinforma* 19: 629–640
- Wang W, Wright EM, Uebersax MA, Cichy K (2022) A pilot-scale dry bean canning and

evaluation protocol. J Food Process Preserv 46: 1–12

Wright D, Kelly (2011) Mapping QTL for seed yield and canning quality following processing of black bean (*Phaseolus vulgaris* L.). Euphytica 179: 471–484

APPENDIX

Table S3.1 QTL identified by genome-wide association analysis over 2018 and 2019 using BLINK and/or FarmCPU models.

Trait	SNP	Chr	Position	P value	year	Model	^a SNP Effect
a*	S02_48372028	2	48372028	1.72E-16	2018	FarmCPU	3.1
	S02_48372028	2	48372028	2.86E-13	2019	FarmCPU-Blink	1.0
App	S02_46696702	2	46696702	9.61E-13	2018	Blink	0.5
	S02_46696702	2	46696702	1.55E-08	2019	Blink	0.4
	S03_52557077	3	52557077	2.49E-10	2018	Blink	0.4
	S03_52994372	3	52994372	1.88E-14	2019	Blink	-0.3
	S05_1300069	5	1300069	8.03E-06	2018	FarmCPU	-0.1
	S05_573193	5	573193	7.86E-06	2019	Blink	0.0
	S09_14033859	9	14033859	1.71E-08	2018	Blink	0.5
	S09_14534494	9	14534494	3.57E-07	2019	FarmCPU	-0.2
b*	S02_48372028	2	48372028	1.52E-15	2018	Blink	6.2
	S02_47632439	2	47632439	4.81E-06	2019	Blink	2.5
Col	S08_7162048	8	7162048	3.09E-09	2018	Blink	-0.3
	S08_7162048	8	7162048	5.07E-09	2019	Blink	-0.6
	S11_53107257	11	53107257	2E-18	2018	Blink	0.6
	S11_53107257	11	53107257	4.46E-19	2019	Blink	1.0
DF	S04_44964231	4	44964231	9.23E-07	2018	Blink	-0.9
	S04_43411364	4	43411364	1.29E-06	2019	Blink	-1.1
	S05_30486494	5	30486494	7.24E-08	2018	FarmCPU	6.6
	S05_31057655	5	31057655	0.0000109	2019	FarmCPU	4.6
DPM	S01_14016092	1	14016092	8.08E-07	2018	Blink	0.4
	S01_14016092	1	14016092	4.08E-07	2019	FarmCPU	-0.6
L*	S02_48372028	2	48372028	1.41E-06	2018	FarmCPU	5.5
	S02_49587954	2	49587954	3.11E-06	2019	Blink	-0.7
SW	S03_12162718	3	12162718	2.81E-06	2018	FarmCPU	-1.1
	S03_12162718	3	12162718	4.49E-19	2019	FarmCPU	-1.8
	S04_44964407	4	44964407	0.0000143	2018	Blink	0.7
	S04_45177713	4	45177713	0.0000132	2019	Blink	1.1
	S06_19125800	6	19125800	8.33E-09	2018	Blink	1.6
	S06_19002691	6	19002691	3.76E-06	2019	FarmCPU	-0.5
Text	S02_38297750	2	38297750	2.05E-14	2018	Blink	-9.5
	S02_38297750	2	38297750	1.12E-12	2019	Blink	-1.9
	S05_39412197	5	39412197	0.0000338	2018	FarmCPU	5.1
	S05_39645120	5	39645120	6.45E-07	2019	FarmCPU	0.2
	S07_9377855	7	9377855	0.0000276	2018	FarmCPU	-1.1
	S07_9377855	7	9377855	3.6E-08	2019	FarmCPU	-6.0
	S10_43229247	10	43229247	0.0000382	2018	FarmCPU	-2.5
	S10_43647153	10	43647153	1.75E-06	2019	FarmCPU-Blink	7.6
YD	S04_1577686	4	1577686	1.49E-08	2018	FarmCPU	69
	S04_1835757	4	1835757	4.14E-06	2019	Blink	101

^aSNP effect is the phenotypic difference between homozygous genotypes.

Table S3.2 Prediction accuracy of agronomic and canning quality traits in the BBL using regularized selection indices (RSI), single-trait (ST) and multi-trait (MT) models in the training population planted in 2018 and 2019.

Trait	Year	h^2	RIS	ST	MT-RSI	MT-Cor [‡]	ST-QTL
YD	2018	0.69	0.16	0.58	0.58	0.61	0.59
YD	2019	0.58	0.06	0.66	0.66	0.66	0.60
SW	2018	0.87	0.09	0.75	0.76	-	0.77
SW	2019	0.88	0.02	0.83	0.83	-	0.78
DF	2018	0.74	0.19	0.44	0.46	-	0.45
DF	2019	0.69	0.00	0.61	0.62	-	0.57
DPM	2018	0.83	0.21	0.56	0.57	-	0.55
DPM	2019	0.26	0.00	0.61	0.62	-	0.32
App	2018	0.83	0.05	0.63	0.63	0.67	0.67
App	2019	0.72	0.06	0.66	0.66	0.69	0.60
Text	2018	0.80	0.12	0.57	0.57	-	0.63
Text	2019	0.94	0.19	0.92	0.91	-	0.89
Col	2018	0.88	0.08	0.75	0.75	-	0.78
Col	2019	0.94	-0.01	0.84	0.84	-	0.87
L*	2018	0.82	0.31	0.80	0.80	-	0.79
L*	2019	0.82	-0.01	0.81	0.81	-	0.79
a*	2018	0.87	0.15	0.75	0.74	-	0.80
a*	2019	0.86	0.00	0.81	0.81	-	0.80
b*	2018	0.90	0.09	0.81	0.80	-	0.85
b*	2019	0.95	0.06	0.87	0.87	-	0.85

Prediction accuracy is expressed as average correlation between observed and predicted genetic value across all 100 partitions in the training data set.

[‡]The MT-Cor model used seed weight and texture as secondary traits to predict seed yield and appearance, respectively.

Table S3.3 Prediction accuracy of agronomic and canning quality traits in the BBL using single-trait (ST) and multi-trait (MT) models in the testing population planted 2019. Data from 2019 in the training population alone and in combination with several proportion of testing set were used to train the models.

Trait	h ²	Trn + 0% Tst					Trn + 10% Tst					Trn + 20% Tst					Trn + 30% Tst					Trn + 40% Tst					Trn + 50% Tst				
		ST	MT-RSI	MT-Cor [‡]	ST-QTL	MT-RCQ ^{‡‡}	ST	MT-RSI	MT-Cor	ST-QTL	MT-RCQ	ST	MT-RSI	MT-Cor	ST-QTL	MT-RCQ	ST	MT-RSI	MT-Cor	ST-QTL	MT-RCQ	ST	MT-RSI	MT-Cor	ST-QTL	MT-RCQ	ST	MT-RSI	MT-Cor	ST-QTL	MT-RCQ
YD	0.58	0.22	0.25	0.30	0.19	0.27	0.27	0.26	0.33	0.26	0.30	0.32	0.31	0.38	0.31	0.35	0.36	0.34	0.40	0.34	0.37	0.37	0.35	0.42	0.35	0.39	0.38	0.36	0.42	0.37	0.39
SW	0.88	0.40	0.45	-	0.56	0.57	0.44	0.44	-	0.57	0.58	0.47	0.47	-	0.60	0.60	0.48	0.49	-	0.61	0.61	0.49	0.49	-	0.63	0.63	0.51	0.52	-	0.64	0.64
DF	0.69	0.32	0.20	-	0.19	0.18	0.37	0.34	-	0.30	0.26	0.43	0.42	-	0.39	0.38	0.45	0.44	-	0.44	0.42	0.49	0.50	-	0.49	0.48	0.52	0.51	-	0.52	0.51
DPM	0.26	0.03	0.09	-	0.08	0.07	0.05	0.07	-	0.06	0.09	0.07	0.08	-	0.09	0.10	0.09	0.09	-	0.10	0.10	0.10	0.10	-	0.11	0.11	0.11	0.11	-	0.13	0.12
App	0.72	0.31	0.38	0.50	0.35	0.46	0.35	0.33	0.48	0.36	0.42	0.35	0.34	0.47	0.37	0.43	0.35	0.34	0.47	0.38	0.45	0.34	0.33	0.46	0.37	0.44	0.33	0.32	0.45	0.37	0.45
Text	0.94	0.42	0.39	-	0.39	0.38	0.47	0.49	-	0.48	0.49	0.50	0.53	-	0.52	0.54	0.51	0.54	-	0.54	0.56	0.54	0.57	-	0.57	0.59	0.54	0.57	-	0.56	0.59
Col	0.94	0.78	0.77	-	0.86	0.85	0.81	0.80	-	0.86	0.86	0.81	0.81	-	0.87	0.86	0.83	0.83	-	0.87	0.87	0.84	0.84	-	0.88	0.88	0.84	0.84	-	0.88	0.88
L*	0.82	0.61	0.63	-	0.63	0.63	0.64	0.64	-	0.64	0.65	0.65	0.65	-	0.65	0.65	0.68	0.68	-	0.68	0.68	0.69	0.69	-	0.69	0.69	0.70	0.70	-	0.70	0.70
a*	0.86	0.74	0.73	-	0.75	0.73	0.73	0.73	-	0.74	0.73	0.73	0.73	-	0.74	0.74	0.74	0.74	-	0.75	0.75	0.74	0.74	-	0.75	0.75	0.74	0.74	-	0.75	0.75
b*	0.95	0.71	0.70	-	0.63	0.60	0.73	0.73	-	0.65	0.65	0.74	0.74	-	0.68	0.68	0.77	0.77	-	0.71	0.71	0.78	0.78	-	0.74	0.73	0.79	0.78	-	0.75	0.75

[‡] The MT-Cor model used seed weight and texture as secondary traits to predict seed yield and appearance, respectively.

^{‡‡} The MT-RCQ model used QTL as fixed variables and secondary traits in the model (RSI and correlated traits (for appearance and seed yield))

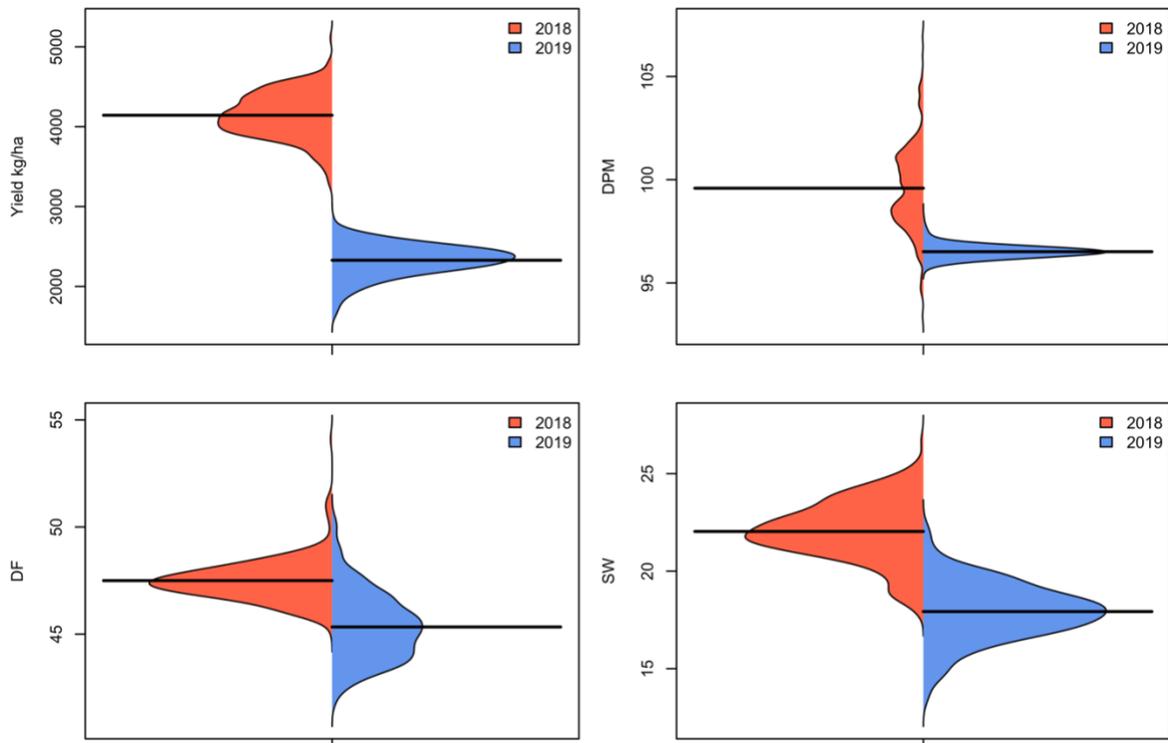


Figure S3.1 Bean plots of yield (YD), days to physiological maturity (DPM), days to flowering (DF), and seed weight (SW) of the BBL. The trials were carried out in 2018 and 2019 in MI. Best linear unbiased estimator (BLUE) were obtained for each trait and trial, adjusting for spatial effects in the field.

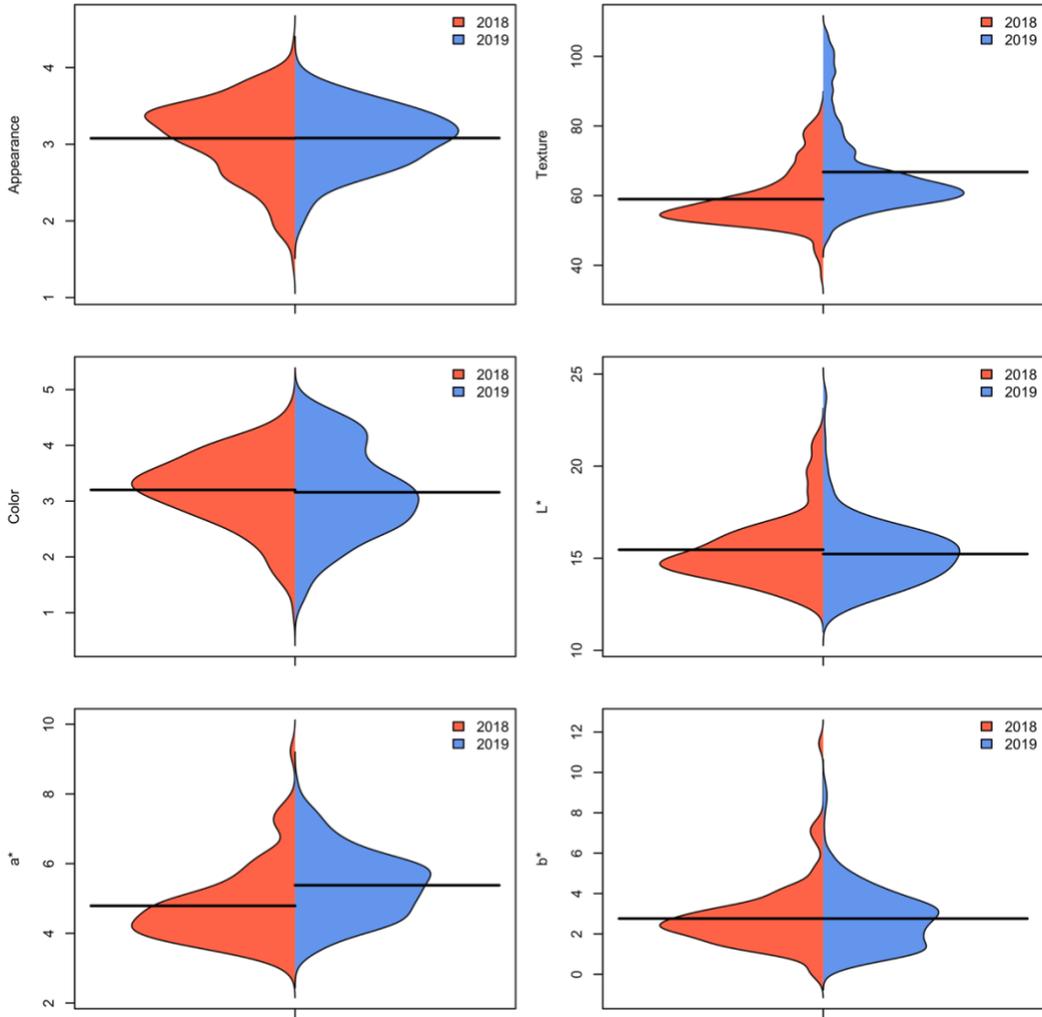


Figure S3.2 Bean plots of appearance (App), texture (Text), color (Col), and CIELAB color space (L^* , a^* , b^*) of the BBL. The trials were carried out in 2018 and 2019 in MI. Best linear unbiased estimator (BLUE) were obtained for each trait and trial, adjusting for spatial effects in the field.

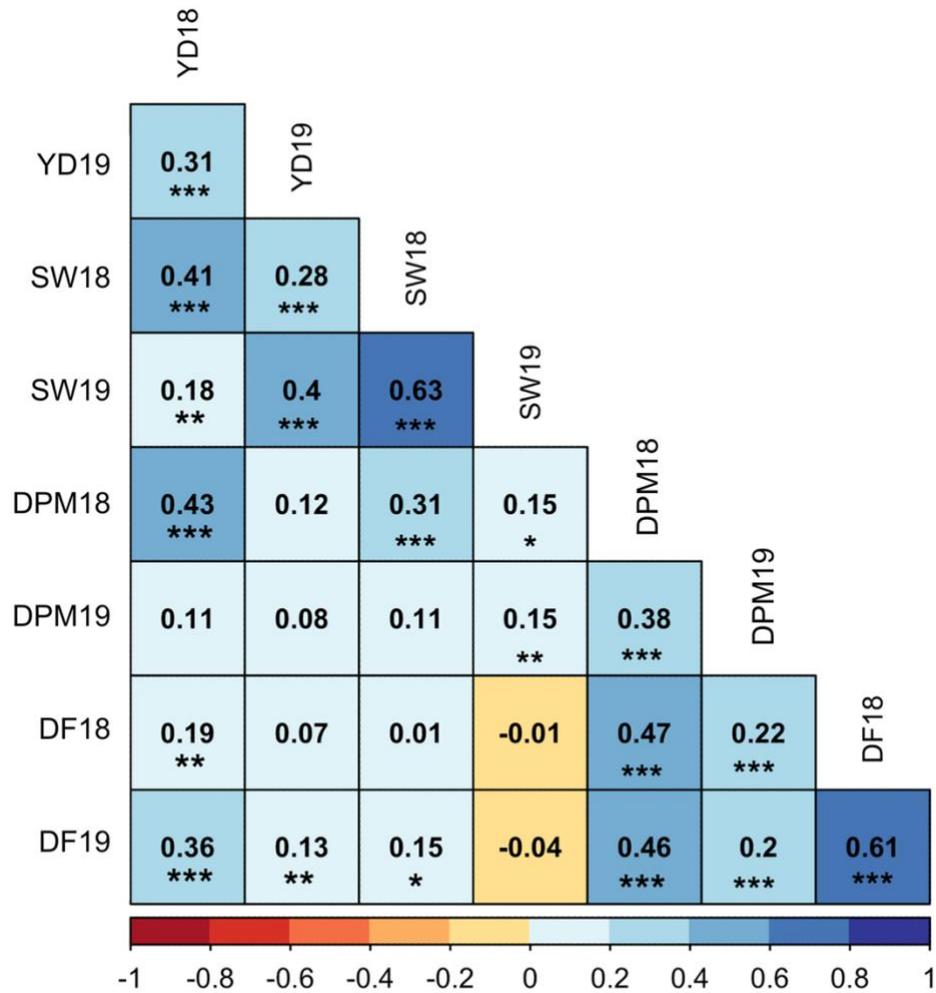


Figure S3.3 Pearson's correlation coefficients between best linear unbiased prediction (BLUP). Significance of correlations indicated as ***: $p < .001$; **: $p < .01$; *: $p < .05$. Yield (YD), seed weight (SW), Days to physiological maturity (DPM), and Days to flowering (DF).

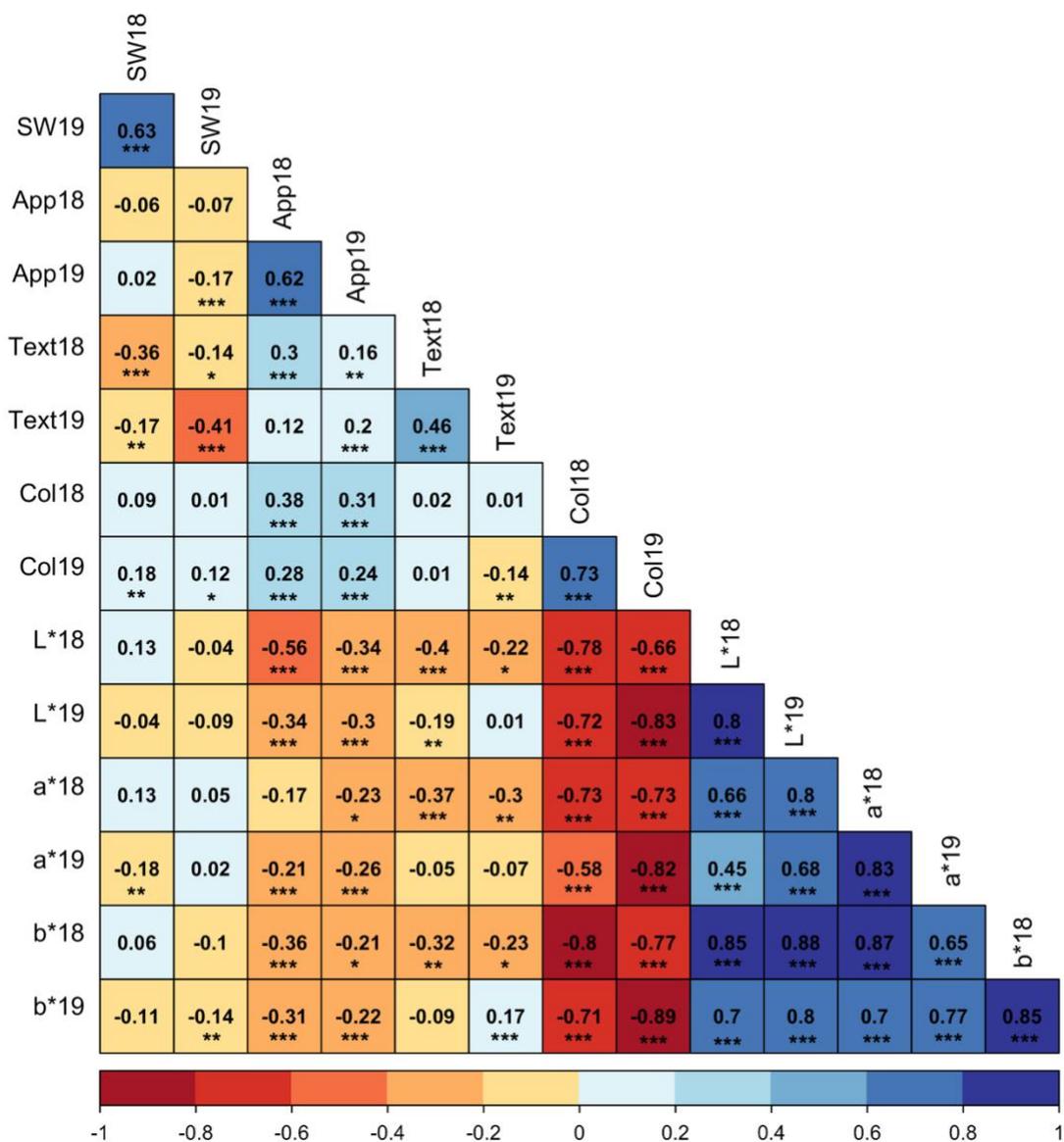


Figure S3.4 Pearson's correlation coefficients between best linear unbiased prediction (BLUP). Significance of correlations indicated as ***: $p < .001$; **: $p < .01$; *: $p < .05$. Seed weight (SW), appearance (App), texture (Text), color (Col), and CIELAB color space (L*, a*, b*).

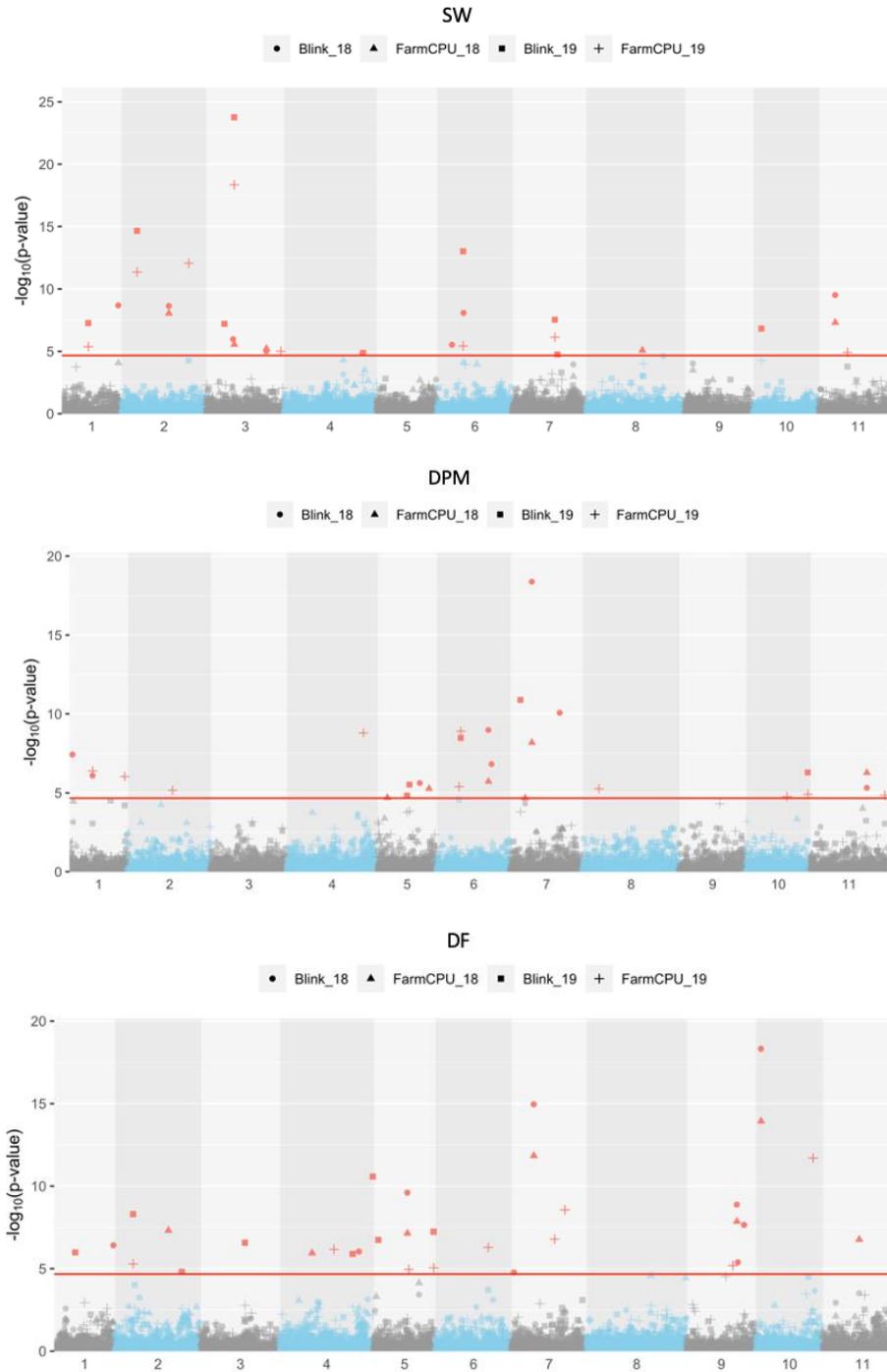


Figure S3.5 Genome-wide association analysis of seed weight (SW), days to physiological maturity (DPM), and days to flowering (DF). The red lines indicate the Bonferroni threshold ($\alpha = 0.05$).

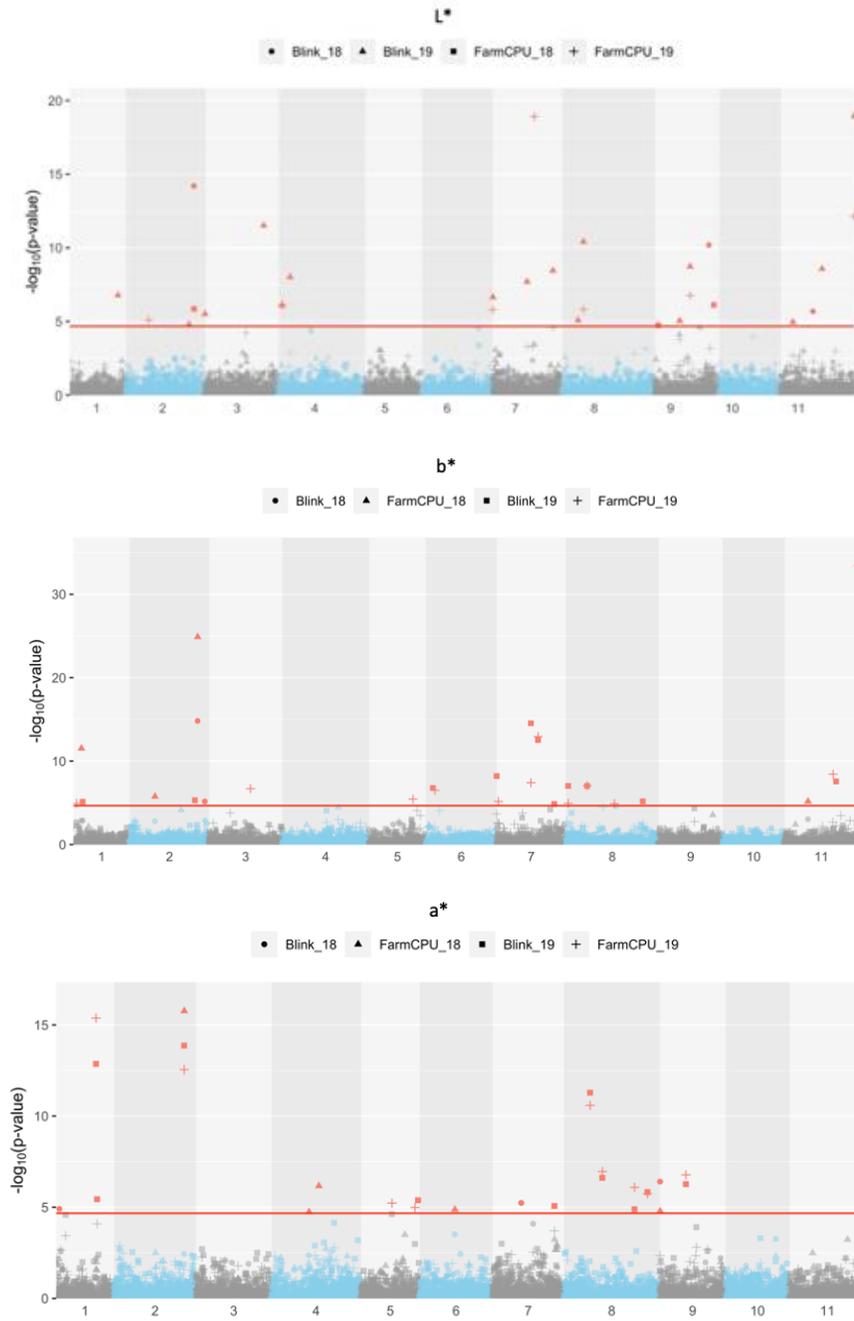


Figure S3.6 Genome-wide association analysis of CIELAB color space (L^* , b^* , a^*). The red lines indicate the Bonferroni threshold ($\alpha = 0.05$).

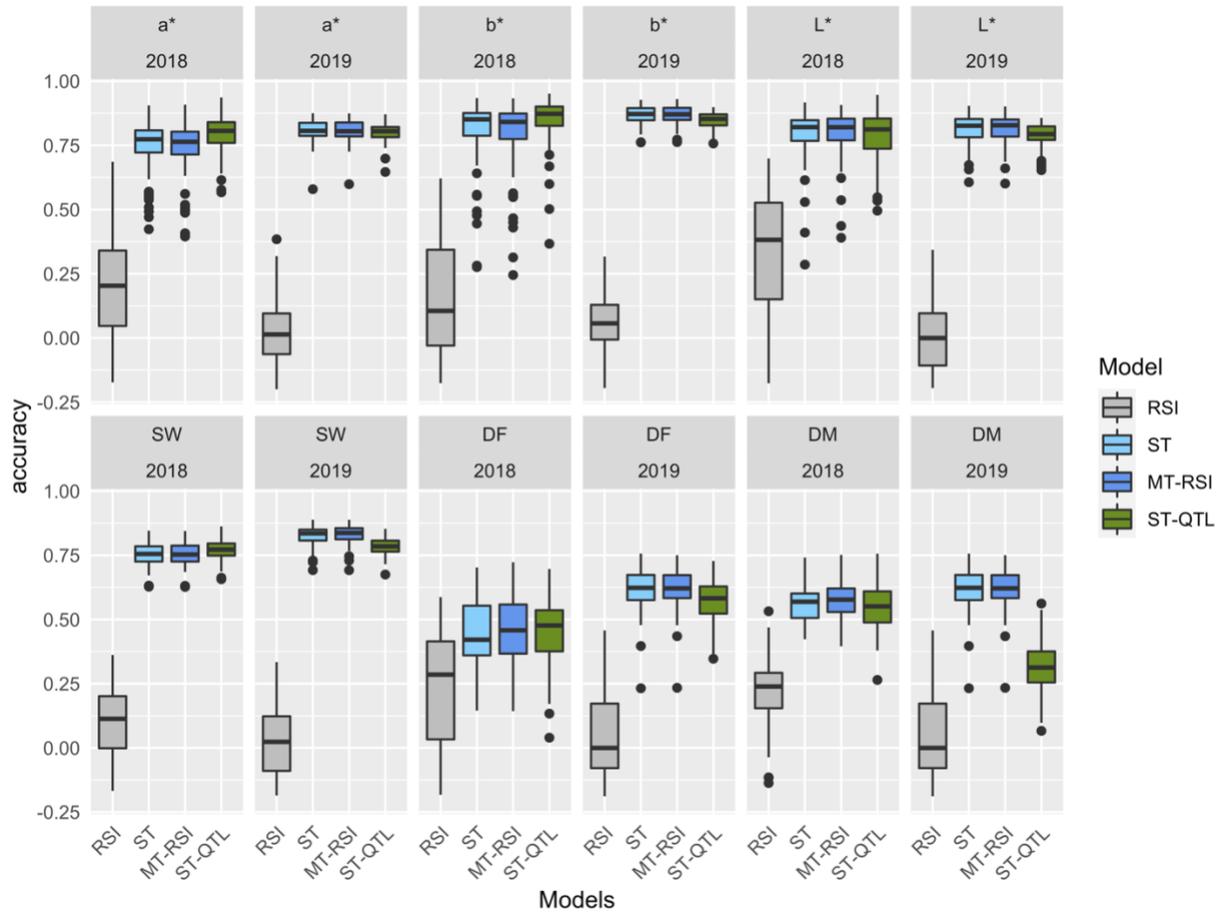


Figure S3.7 Prediction accuracy of CIELAB color space (a^* , b^* , L^*), seed weight (SW), days to flowering (DF), and days to maturity (DPM) in the BBL using RSI, single trait (ST) and multi-trait (MT) models. The distribution of boxplots represents 100 partitions in the training set.

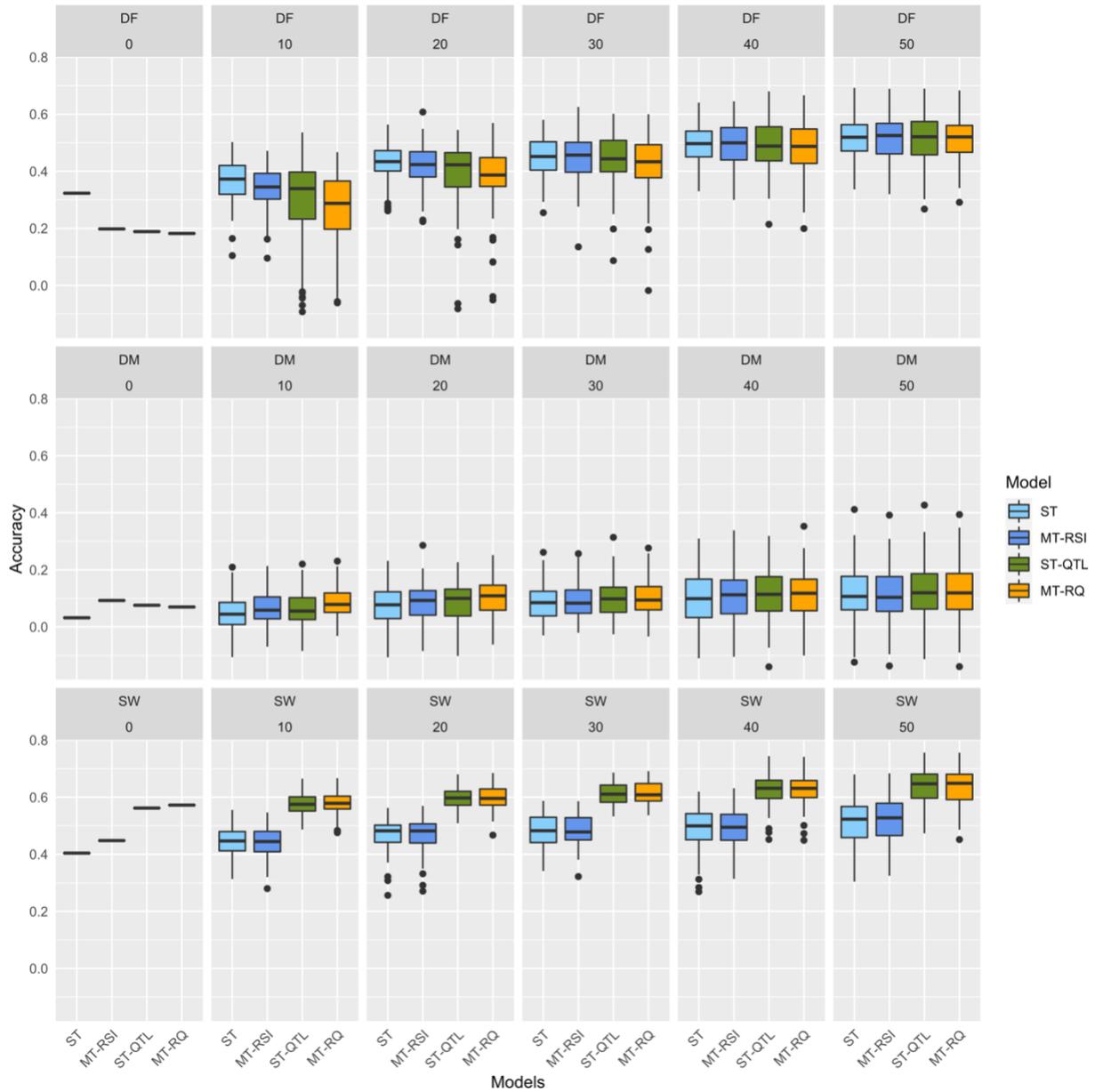


Figure S3.8 Prediction accuracy of seed weight (SW), days to physiological maturity (DPM), and days to flowering (DF) in the 2019 testing set comprised of 140 black breeding lines. A) different proportions of the testing set were included in the models (0%, 10% =14, 20% = 28, 30%=42, 40%= 56, 50%=70). The distribution of boxplots represents 100-fold cross-validations.

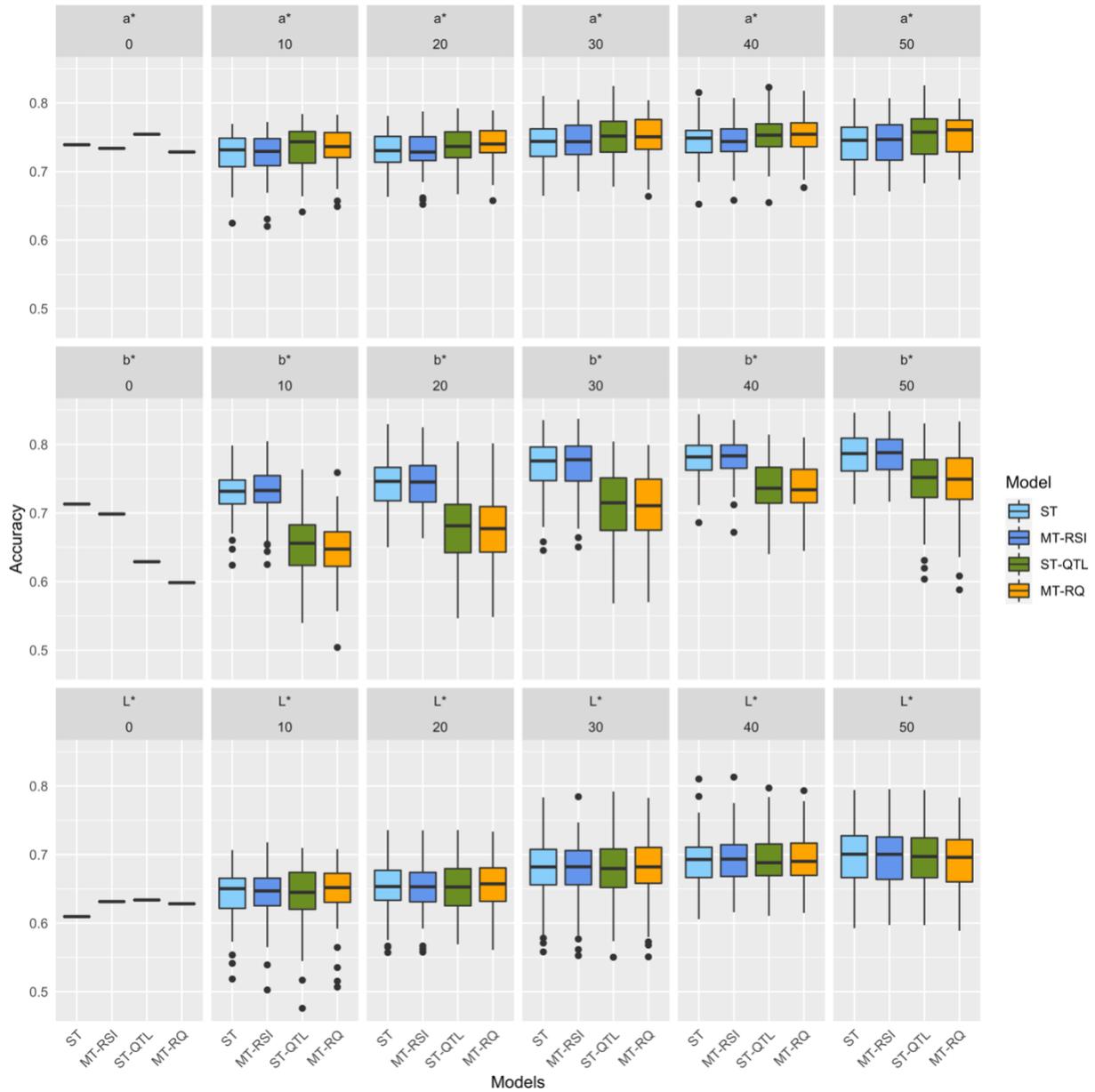


Figure S3.9 Prediction accuracy of CIELAB color space (a^* , b^* , L^*) in the 2019 testing set comprised of 140 black breeding lines. A) different proportions of the testing set were included in the models (0%, 10% =14, 20% = 28, 30%=42, 40%= 56, 50%=70). The distribution of boxplots represents 100-fold cross-validations.

**CHAPTER 4: GENOME-WIDE ASSOCIATION AND GENOMIC PREDICTION FOR
FE-ZN CONCENTRATION AND FE BIOAVAILABILITY IN A YELLOW BEAN
COLLECTION OF DRY BEANS**

Abstract

Dry bean is a nutrient-dense food targeted in biofortification programs to increase seed iron and zinc levels. An assumption of breeding for increased mineral content is that more iron and zinc will be available to humans consuming the biofortified food. However, mineral bioavailability is influenced not just by the mineral content but by other seed compounds the minerals interact with, such as polyphenols and phytate. This study characterized a diversity panel of 295 genotypes comprising the Yellow Bean Collection (YBC) for seed Fe and Zn concentration, Fe bioavailability (FeBio), and seed yield over two years and two field locations. The genetic architecture of each trait was elucidated via genome-wide association (GWA) and genomic prediction (GP) was evaluated using the reproducing kernel hilbert space (RKHS) and the sparse selection index (SSI) method. Additionally, 82 yellow breeding lines were characterized for seed Fe and Zn concentration and seed yield and were used as prediction set in GP models. Large phenotypic variability was identified in all traits evaluated, and variations of up to 2.8 and 13.7-fold were observed for Fe concentration and FeBio, respectively. Prediction accuracies in the YBC ranged from a low of 0.1 for Fe concentration and a high of 0.8 for FeBio, and an accuracy improvement of 0.1 was observed when QTL identified through GWA were used as fixed effects for FeBio. In the prediction set, SSI led to 1%–6% increases in accuracy, relative to the RKHS. This study provides evidence of the lack of correlation between FeBio estimated in vitro and Fe concentration, and reveals the high accuracy of GP for FeBio, which can reduce the high cost of measuring this trait.

Introduction

Dry bean is the most important legume for human consumption worldwide (FAO 2022), providing high levels of protein, dietary fiber and micronutrients such as Fe and Zn (Uebersax et al., 2022). Biofortification initiatives since 2005 have focused on increasing dry bean Fe and Zn concentrations through breeding, as a means to improve the nutritional status of humans suffering from Fe and Zn deficiencies (Beebe, 2020). Dry beans are also an appealing crop choice for biofortification because of their low environmental footprint, ability for symbiotic nitrogen fixation, and long shelf life that minimizes food waste (Willett et al., 2019).

As bean breeders have worked for the last two decades to biofortify beans by increasing minerals concentrations, there has also been major efforts to understand the genetic architecture of seed Fe and Zn accumulation (Blair et al., 2011, 2013; Blair & Izquierdo, 2012; K. Cichy et al.,

2022; K. A. Cichy et al., 2009). A meta-QTL (MQTL) analysis identified 12 stable MQTLs over different genetic backgrounds and environments (Izquierdo et al., 2018). In addition to breeding for increased seed micronutrient concentrations, the importance of breeding for improved Fe bioavailability has also been identified in multiple studies indicating Fe concentration and Fe bioavailability are not necessarily positively correlated with each other (Glahn et al., 2020; Katuuramu et al., 2018, 2021). Increased Fe levels are often associated with increased amounts of polyphenols and phytate in the seed and both compounds form complexes with Fe that can pass through the human gut undigested, thereby negating the potential benefit of the higher Fe concentration (Tako et al., 2015). The polyphenol profile varies among seed coat colors (Tako et al., 2014), and darker seed coats tend to have a more diverse array of polyphenols that reduce the Fe absorption (Wiesinger et al., 2019). Cooking time is also correlated with Fe bioavailability, which may relate to the tendency for fast-cooking genotypes to have lighter colored seed coats, thereby also having less Fe inhibitory polyphenols (Wiesinger et al., 2018).

Seed Fe and Zn concentrations are quantitative traits controlled for many loci across the genome in dry beans (Izquierdo et al., 2018), and are strongly influenced by the environment (Katuuramu et al., 2021). Although Fe bioavailability is also a quantitative trait, a study reported across nine locations in Uganda found it to be stable across environments (Katuuramu et al., 2021). Fe bioavailability is measured via an *in vitro* Caco-2 cell culture assay, and while it is faster than animal studies, it is a trait that is out of reach of most breeding programs (Glahn et al., 1998). One study in common bean has dissected the genetic architecture of Fe bioavailability through genome-wide association (GWA), identifying five SNP associations distributed on chromosomes 6, 7, and 11 with phenotypic variability explained by the associated markers ranging from 8% to 13% (Katuuramu et al., 2018).

Genomic prediction uses the genotype and phenotype of training datasets to estimate the phenotypes of new lines (testing dataset), reducing the phenotyping and increasing selection intensity (Bernardo, 2020). In genomic selection, the phenotypes are predicted from all genetic markers together (Meuwissen et al., 2001), and the accuracy represented as the Pearson correlation of observed vs predicted phenotypes, depends on the heritability of the trait, linkage disequilibrium (LD) with causal loci, population size and the genetic relationship between the training and testing datasets (Voss-Fels et al., 2019). The composition of the training dataset and its relationship with the new lines to be predicted are essential to maximize prediction accuracy (de los Campos et al.,

2013; Isidro y Sánchez & Akdemir, 2021). Although several approaches have been proposed to optimize the training population, most of them have the assumption that one training is optimal for all individuals in the testing dataset (Lopez-Cruz et al., 2022). Taking into consideration that a high level of genetic heterogeneity is plausible in breeding programs, the use of only one optimal population may include individuals distantly related to the individuals in the testing dataset, reducing predictive ability (Lorenz & Smith, 2015). Recently, a sparse selection index (SSI) was proposed to identify a training set for each individual in the testing set (Lopez-Cruz & De Los Campos, 2021), and the use of this approach has increased prediction ability in multigeneration data up to 10% and 17% in wheat maize, respectively, compared to genomic best linear unbiased predictor (GBLUP) (Lopez-Cruz et al., 2021, 2022).

Genetic variability for Fe and Zn concentration and Fe bioavailability is abundant, both among and within market class (Katuuramu et al., 2021). However, lighter seed types such as pale yellow beans have been found to be high in bioavailable Fe (Wiesinger et al., 2018). Furthermore, pale yellow beans are associated with fast-cooking times and palatability (Birachi et al., 2020; K. A. Cichy et al., 2015). The objectives of this research were to *i*) evaluate the Fe and Zn concentration and Fe bioavailability of a previously developed Yellow Bean Collection (YBC) (Sadohara et al., 2022) *ii*) identify genomic regions that are associated with Fe and Zn concentrations and bioavailability via GWA, and *iii*) estimate the prediction ability of genomic prediction for these traits.

Materials and methods

Plant material

The yellow bean collection (YBC) comprised of 295 *Phaseolus vulgaris* L. accessions was grown at the Michigan State University Montcalm Research Farm in Entrican, MI, in 2018 and 2019, and at University of Nebraska field sites in Scottsbluff and Mitchell, Nebraska, in 2018 and 2019, respectively. In all years and locations the YBC was grown in a randomized complete block design with two replications as described by Sadohara et al. (2022). Additionally, 82 F5 yellow bean breeding lines were planted at the Montcalm Research Farm in 2019. These 82 breeding lines were generated using seven YBC accessions as parents in various biparental crosses. The YBC planted in both years were used as the training set, while the 82 lines evaluated in 2019 were used as the prediction set to simulate the implementation of GP using a pre-breeding population. The local standard agricultural practices were followed for research plot scale dry bean production

(Sadohara et al., 2022). Following harvest, seed weight (SW) and seed yield (YD) were collected. SW (g) was obtained from weighing 100 seeds and YD (kg ha⁻¹) was calculated based on the plot size and corrected to the moisture content in the seed of 18%.

Mineral concentration (Michigan and Nebraska)

The protocol described by Glahn et al. (2020) was used to measure Fe and Zn from raw seeds in the YBC grown in MI and NE in 2018 and 2019 and the breeding lines grown in MI in 2019. Briefly, seeds were rinsed and cleaned with distilled water to remove dust and debris. The cleaned seeds were lyophilized and milled, and 0.50 g of the powder was predigested in boro-silicate glass tubes with concentrated ultrapure nitric acid and perchloric acid mixture (60:40 v/v) for 16 h at room temperature. Samples were then placed in a digestion block and heated incrementally over 4 h to a temperature of 120°C with refluxing. After incubating at 120°C, ultrapure nitric acid was subsequently added to each sample before raising the digestion block temperature to 145°C. Digested samples were resuspended and then raised to 190°C and maintained in ultrapure water before analysis using ICP-AES (inductively coupled plasma atomic emission spectrometry; Thermo iCAP 6500 Series, Thermo Scientific) with quality control standards following every 10 samples.

Fe bioavailability (Michigan only)

The YBC grown in MI in 2018 and 2019 was cooked using an automated Mattson cooker as reported previously (Sadohara et al., 2022). The cooked samples were lyophilized and milled and 0.50-g of powder was subjected to an in vitro digestion/Caco-2 cell culture model for the determination of Fe bioavailability, as described previously by Glahn et al. (1998). Fe uptake is measured as the increase in Caco-2 cell ferritin production (ng ferritin per milligram of total cell protein) following simulated gastric and intestinal digestion. Fe bioavailability is expressed as a percentage score of Caco-2 cell ferritin formation that is relative to a control cooked/lyophilized/milled white kidney bean (Snowdon). The kidney bean control was run with each assay to index the ferritin/total cell protein ratios of the Caco-2 cells over the course of experimentation. The Snowdon kidney bean was used as a control since this cultivar was reported with high Fe bioavailability and low polyphenol concentration that inhibit the absorption of iron (Wiesinger et al., 2019). The mean ferritin formation values of Snowdon across assays were 19.71 and 15.78 ng ferritin/mg total cell protein in 2018 and 2019, respectively. Fe bioavailability was not measured in the 82 breeding lines.

Statistical analyses

The rows and columns from the field were used as random effects to fit a linear mixed model for SW and YD using the functions “SpATS” and “SAP” for the R package SpATS (Rodríguez-Álvarez et al., 2018). The effect of the genotype on the phenotype was fitted as fixed to obtain the best linear unbiased estimator (BLUE) of SW and YD. The mineral concentrations were collected in one field replication in Michigan and Nebraska during two growing seasons (2018–2019), while Fe bioavailability was collected in one field replication only in Michigan during two growing seasons (2018–2019). The mean of two technical replications were used as the mineral concentration and Fe bioavailability for each environment and year. The analysis of variance (ANOVA) was conducted using the R package statgenGxE using a factorial structure of locations per year (Van, 2022). The mixed model used year, location and location:year terms as fixed, while the effect of genotype, genotype:year, and genotype:location was treated as random.

Genotypic data

DNA was extracted from trifoliolate leaves using NucleoSpin Plant II Kit (Macherey–Nagel, Duren, Germany) following the ‘Genomic DNA from plant’s protocol as described previously by Sadohara et al. (2022). DNA concentration was measured using Quant-iT™ PicoGreen™ dsDNA Assay Kit (Invitrogen, Waltham, MA, USA), and 10 ng/μL of DNA was used for GBS library preparation according to Elshire et al. (2011) with a single restriction enzyme, ApeKI. Each plate of 96-wells was sequenced in a lane of an Illumina HiSeq platform using single-end reads. The libraries were demultiplexed using NGSEP (v3.1.2) (Tello et al., 2019). Adapters and low-quality bases from the raw sequencing data were trimmed using Cutadapt v 1.16, and the processed reads were aligned to the reference genome of *P. vulgaris* v2.1 G19833 (Schmutz et al., 2014) using Bowtie2 (v2.2.30) (Langmead & Salzberg, 2012) with default parameters. The SNP calling was carried out by using NGSEP software following the recommended parameters for GBS data (Perea et al., 2016). The merged genotypic matrix was filtered with NGSEP for variants that were in the predicted repetitive regions of the common bean genome reported by Lobaton et al. (2018), non-biallelic, genotype quality above 30, a maximum observed heterozygosity of 0.05 per SNP, more than 50% of missing data per site, and minor allele frequency (MAF) above 0.05. Besides, SNPs in linkage disequilibrium above 0.9 using a window of 500 SNPs were removed using Bcftools (Li, 2011). The resulting genotype matrix was imputed using Beagle V5.4 (Browning et al., 2018).

Genome-wide association study (GWAS)

GWAS was conducted using the multi-locus methods Bayesian-information and linkage-disequilibrium iteratively nested keyway (BLINK) and Fixed and random model Circulating Probability Unification (FarmCPU) implemented in the R package GAPIT v3 (Wang & Zhang, 2021). For the GWAS, the complete set of samples of each year and location was used, and a Bonferroni threshold of $\alpha=0.05$ was used to identify associations. When an association was found using the Bonferroni correction, a false discovery rate (FDR) of $\alpha=0.05$ was used to determine whether the same association was present in any other year-location. The QTL associations were defined as consistent when presented in more than one year-location with the same SNP, or when two different SNPs were associated in the same genomic region between years, and the distance between them was not greater than 100 kb.

Genomic prediction

In total, three prediction models were assessed: reproducing kernel hilbert space (RKHS), sparse selection indices (SSI), and SSI using consistent QTL identified through GWAS as fixed variables. The SSI was obtained by imposing an L1-penalty on a selection index using additive genomic relationships (GSSI). To assess the prediction ability of the models, the training set was randomly divided into 70% and 30% to train and validate the models, respectively. To optimize the penalization value λ in SSI, 10-fold cross validation was carried out in the training subset, and SSI were derived over a grid of 100 values of λ . The accuracy measured as the Pearson correlation between SSI and each trait was used to identify the value of λ that maximized accuracy in each cross validation. The optimal value of λ was defined as the average value of λ that maximized accuracy across each cross-validation and was used to predict the validation subset. The training-validation procedure described above for RKHS and SSI was repeated 100 times by randomly assigning samples from the training set into training and validation subsets, and was implemented using the R package SFSI (Lopez-Cruz et al., 2020). The prediction ability for all models is expressed as a Pearson correlation coefficient between the observed and predicted values in the validation subset.

Prediction set

To simulate the implementation of GP, 82 breeding lines from 7 bi-parental crosses were included in 2019 and used as the prediction set as described previously. RKHS and SSI models were assessed in the prediction set. Additionally, different proportions of individuals of each bi-

parental family (0%, 10%, 20%, 30%) were randomly assigned to the training set to train the models. The procedure described above was repeated 100 times, and was implemented using the R package SFSI (Lopez-Cruz et al., 2020)

Results

Phenotypic data

The YBC was grown in field locations in MI and NE in 2018 and 2019 and was evaluated for seed yield, seed weight, Fe and Zn concentration in MI and NE 2018–2019, and FeBio was evaluated in Michigan 2018–2019. The YBC exhibit large variability for agronomic and nutritional quality traits (Table 1). For all traits except FeBio, the distribution of the phenotypes was normally distributed across locations and years (**Supplementary Figure 4.1–2**). The FeBio distribution was binomial in the two growing seasons measured in Michigan. The concentration of Fe ranged from 43 to 118 ug/g in Michigan 2018, 39 to 106 ug/g in Michigan 2019, 28 to 76 ug/g in Nebraska 2018, and 36 to 72 ug/g in Nebraska 2019 (**Table 4.1**). Four genotypes presented concentrations above 100 ug/g in Michigan, three in 2018 (YBC050, YBC136, and YBC279), and one in 2019 (YBC171). The concentration of Zn ranged from 20 to 41 ug/g in Michigan 2018, 17 to 41 ug/g in Michigan 2019, 19 to 37 ug/g in Nebraska 2018, and 21 to 43 ug/g in Nebraska 2019 (**Table 4.1**). FeBio values in Michigan varied from 11 to 151% in 2018 and from 13 to 118% in 2019, compared to the control Snowdon (**Table 4.1**). In total, twenty-seven genotypes had higher FeBio values than the high FeBio check variety, Snowdon, in both 2018 and 2019.

The Analysis of variance revealed a significant effect of genotype on the agronomic and nutritional quality traits assessed in the current study (**Supplementary Table 4.1**). The most important source of variation for Fe concentration was the location followed for the Year:Location, while genotype explained most of the variation in Zn concentration and FeBio. Indeed, the genotype explained 93%, 5%, and 38% of the phenotypic variation in FeBio, Fe and Zn concentration, respectively. The large effect of the genotype on the variability of FeBio is supported by the high h^2 of FeBio found in 2018 (0.85) and 2019 (0.91).

Despite the differences between location and years, the measurements of all traits were positively correlated across years (**Figure 4.1**). The correlation between Fe concentration and yield was not significant in any environment ranging from -0.05 in Michigan 2018 to 0.10 in Nebraska 2019, while the correlation between yield and Zn concentration was negative in both locations and years. Fe and Zn concentrations were positively correlated, ranging from $r=0.32$ ($p < 0.001$) in

Nebraska in 2018 to $r= 0.58$ ($p < 0.001$) in Michigan in 2019. Both Fe and Zn concentrations were negatively correlated with seed weight in both locations and years. The most stable trait across years (2018–2019) was FeBio, with a correlation between years of 0.93 ($p < 0.001$) in Michigan. FeBio was negatively correlated with yield in 2018 ($r= -0.13$, $p=0.05$), and 2019 ($r= -0.31$, $p=0.001$).

Table 4.1 The mean and the range of yield, seed weight, Fe bioavailability, Fe and Zn concentration of the YBC grown in MI and NE in 2018 and 2019.

Trait	MI 2018		MI 2019		NE 2018		NE 2019	
	Mean	Range	Mean	Range	Mean	Range	Mean	Range
Yield (kg/ha)	2111	1171 - 3299	1654	536 - 3065	1974	653 - 5275	686	52 - 2124
SW (g)	38.0	17.8 - 60.2	39.5	18.4 - 72.8	34.9	14.6 - 54.3	29.7	25.5 - 34.7
Fe ($\mu\text{g/g}$)	69.1	42.8 - 117.7	63.8	38.6 - 105.8	44.7	27.7 - 75.9	51.3	36.3 - 71.7
Zn ($\mu\text{g/g}$)	28.8	20.1 - 41.4	27.7	17 - 41	25.3	18.4 - 36.7	27.0	20.8 - 43.2
FeBio (% of control)	53.5	11 - 151	51.5	12.5 - 118	-	-	-	-

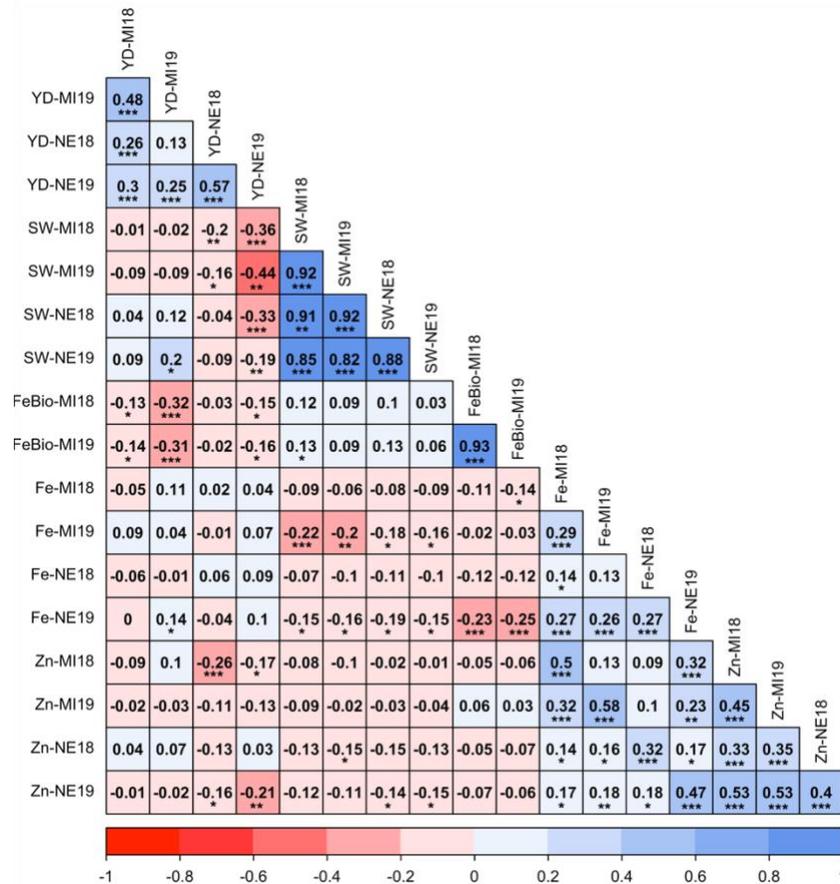


Figure 4.1 Pearson's correlation coefficients between agronomic traits (Seed weight (SW), yield (YD), mineral concentration (Fe, Zn), and Fe bioavailability (FeBio)). Significance of correlations indicated as ***: $p < .001$; **: $p < .01$; *: $p < .05$.

Population structure

In total, 18,357 SNP markers were retained after filtering in the 295 YBC and 82 breeding lines. The genomic relation matrix identified two groups with stronger relationships representing the germplasm of the two gene pools in *P. vulgaris* (**Figure 4.2 A**). A complete description of the genetic diversity and origin of each genotype in the YBC was reported by Sadohara et al. (2022). As expected, the breeding lines included in 2019 were clustered in the same gene pool of their founders (Andean). Genetic diversity was evaluated through principal component analysis (PCA).

The first two principal components explained 53.1% of the variance. PC1 separated the Andean and Middle American gene pools as reported by Sadohara et al. 2022. However, the variance explained by PC1 (48.6%) in this study differ from the PC1 (63.8%) reported by Sadohara et al. (2022). This difference is due different filters applied to the genotype matrix, and the addition of the 82 breeding lines in the present study (**Figure 4.2 B**).

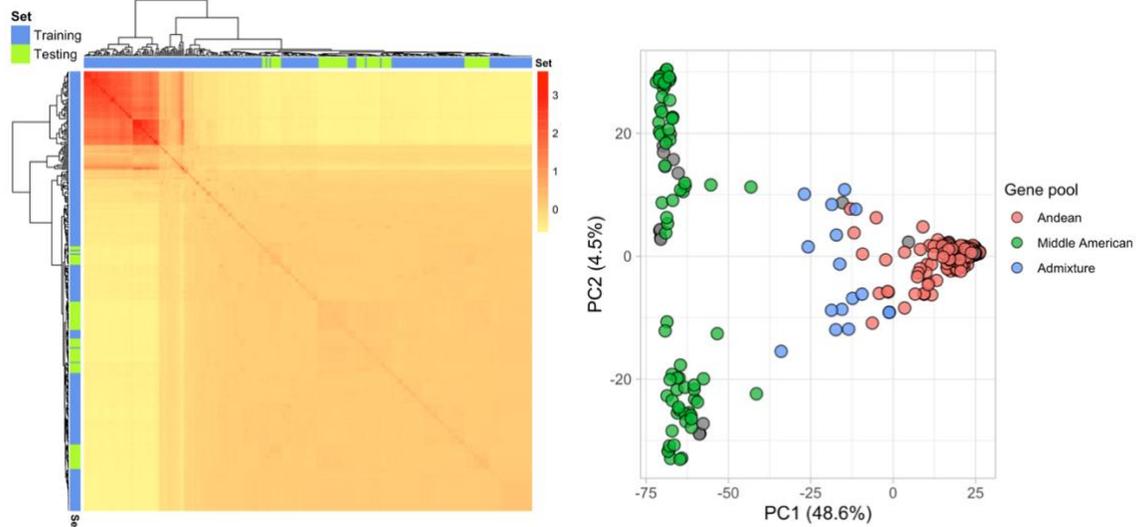


Figure 4.2 Genetic structure of the YBC and advanced yellow RILs with 18,357 SNPs. A Heatmap of the genomic relationship matrix. The blue and green color represent the genotypes in the training and prediction sets, respectively. B The principal component analysis shows the location of each genotype defined by the eigenvectors of the first and second principal components. The color represents the gene pool determined by Sadohara et al. 2022.

Genome-wide association (GWA)

Marker-trait associations were evaluated using the multi-marker approaches FarmCPU and BLINK. In total, 51 SNP associations were identified with significance greater than that established by the Bonferroni correction (2.72×10^{-6}) (**Figure 4.3, Supplementary table 4.2 and Supplementary figure 4.3**). When an association was found using the Bonferroni correction in

only one location or year, the false discovery rate (FDR) of $\alpha=0.05$ was used to determine whether the same association (or association with another SNP located no farther of 100 kb) was present in any other location or year. In total, 3 SNP makers using the FDR threshold supported regions identified using the Bonferroni correction.

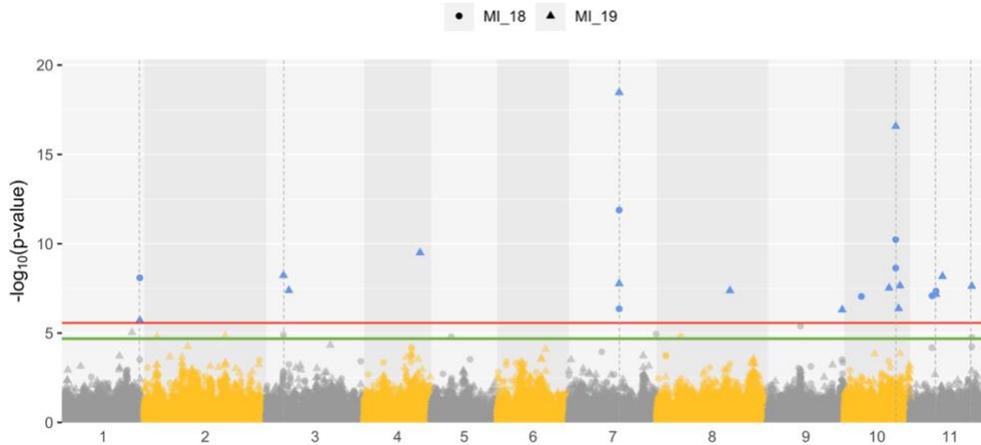


Figure 4.3 Genome-wide association analysis of Fe bioavailability (FeBio). The red and green lines indicate the Bonferroni and FDR threshold at $\alpha = 0.05$, respectively.

In total, six genomic regions were consistently associated with FeBio in Michigan across years, while no consistent associations were identified for Fe and Zn concentration or yield (**Table 4.2**). The SNP effect (representing as the difference between homozygous genotypes) of the association on chromosome 7 for FeBio (% of control) was 45.2 and 46.9 in 2018 and 2019, respectively (**Figure 4.4, Table 4.2**).

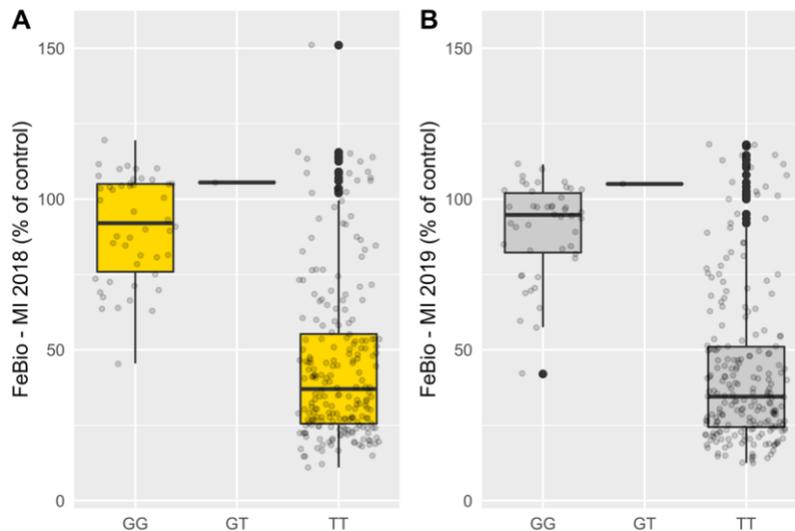


Figure 4.4 Phenotypic effect of SNP at chromosome 7 (29,2 Mb) associated with FeBio in 2018 (A) and 2019 (B).

Table 4.2 QTL for FeBio identified by genome-wide association analysis over years using BLINK and/or FarmCPU models.

Trait	Chr	Position	P.value	Year	Loc	Model	SNP Effect
FeBio	1	51,031,345	8.11E-09	2018	MI	Blink	3.3
	1	51,031,345	1.92E-06	2019	MI	FarmCPU	3.9
	3	2,637,502	1.25E-05	2018	MI	FarmCPU*	21.8
	3	2,649,825	5.91E-09	2019	MI	Blink	20.4
	7	29,169,848	4.43E-07	2018	MI	FarmCPU-Blink	45.2
	7	29,169,848	1.75E-08	2019	MI	FarmCPU-Blink	46.9
	10	42,422,135	5.90E-11	2018	MI	FarmCPU-Blink	3.3
	10	42,422,135	2.67E-17	2019	MI	FarmCPU	3.9
	11	3,626,904	4.75E-08	2018	MI	FarmCPU-Blink	9.3
	11	3,626,904	6.77E-08	2019	MI	FarmCPU	8.0
	11	49,383,743	1.72E-05	2018	MI	FarmCPU*	11.7
	11	49,383,743	2.37E-08	2019	MI	FarmCPU	15.3

*SNPs that are declared associated using FDR < 0.05.

Genomic prediction – training set

In total, 100 partitions were used to assess the prediction accuracy for the Reproducing Kernel Hilbert Spaces (RKHS) and sparse selection indices (SSI) models in the training set composed of 295 YBC. The prediction accuracies of RKHS and SSI models for Fe and Zn concentrations and seed yield are presented in **Figure 4.5 and Supplementary Table 4.3**. The prediction accuracy was affected by trait architecture, and no significant differences were observed across models. RKHS models used all the genotypes in the training subset to train the model, while SSI used a penalization to select the individuals in the training subset to maximize the prediction ability. The average number across locations and years of individuals that support predictions using the SSI ranged from 52% to 71% of individuals in the training subset from Fe concentration and yield, respectively (**Supplementary Table 4.3**). The lowest prediction accuracy found in this study were detected in Fe concentration in Nebraska 2018. A soil analysis revealed that the pH in this location was 8.0. The h^2 (0.05, 0.20) and prediction ability (0.11, 0.14) for Fe and Zn concentration were affected for the alkaline soil of this location. Average prediction accuracy in 2018 and 2019 for yield ranged from 0.33 to 0.65 in Michigan, and 0.51 to 0.66 in Nebraska, respectively (**Figure 4.5**).

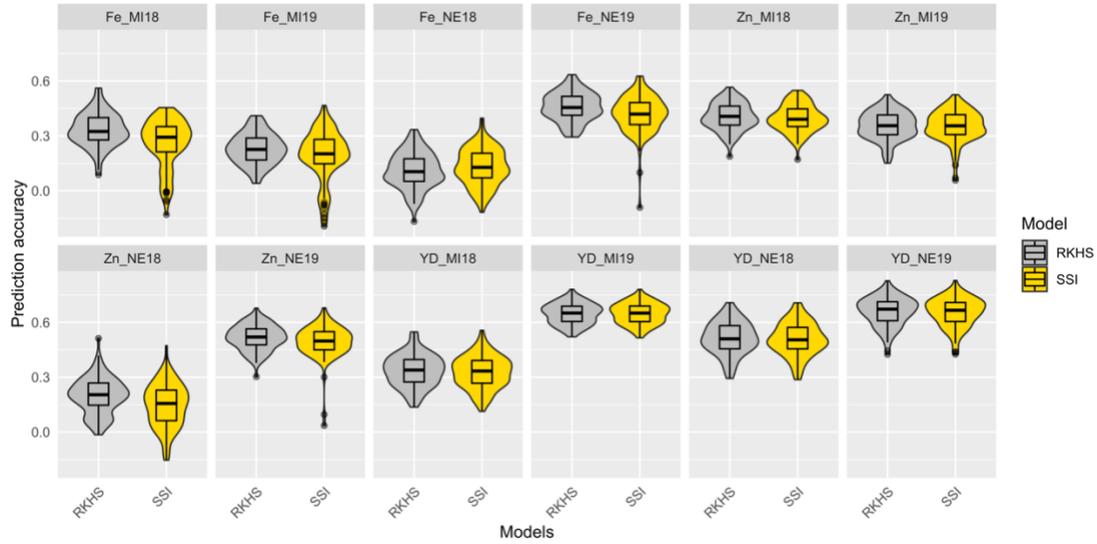


Figure 4.5 Prediction accuracy in the training data set of Fe-Zn concentration and yield using RKHS and SSI. The distribution of boxplots represents 100-fold cross-validations in the training set.

Using the 100 partitions described above, the prediction accuracy for RKHS and SSI models were assessed for FeBio prediction adding the six consistent QTL identified through GWAS (Table 1). Although no differences were observed between RKHS and SSI models without QTL information, an increase in prediction ability was detected when QTL were added to the RKHS model, while the inclusion of QTL in the SSI model negatively impacted the predictions (**Figure 4.6, Supplementary Table 4.3**).

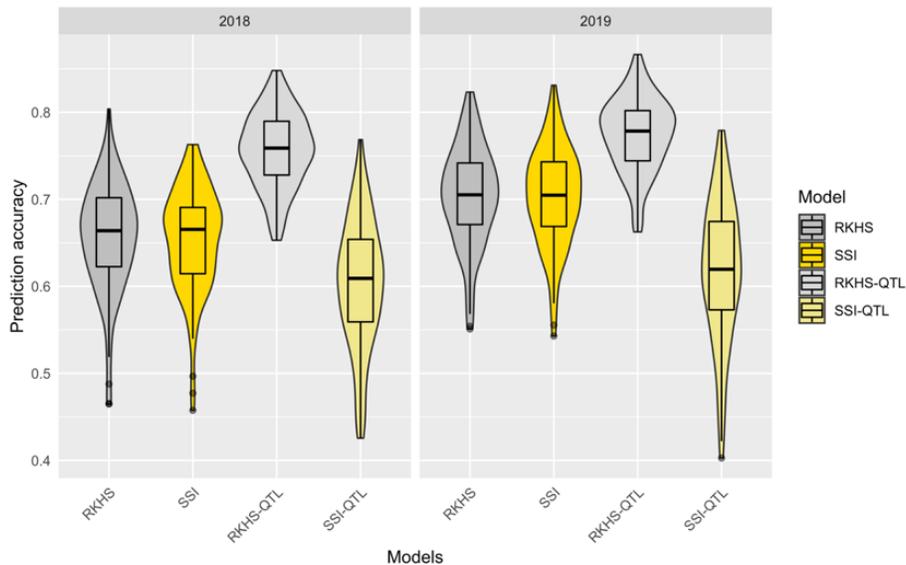


Figure 4.6 Prediction accuracy in the training data set of FeBio using RKHS, SSI and QTL as fixed effects. The distribution of boxplots represents 100-fold cross-validations in the training set.

Genomic prediction – prediction set

To simulate the implementation of GP with a diversity panel where some lines were used in breeding, a set of 82 breeding lines were grown in Michigan 2019 and used as a prediction set. The prediction accuracy of RKHS and SSI models adding different proportions (from 0 to 30%) of the prediction set to train the models are presented in **Figure 4.7 and Supplementary Table 4.4** For the traits measured in the prediction set (seed yield and Fe and Zn concentrations), the prediction accuracy of Fe concentration was not improved when breeding lines were added to train the model. However, the accuracy tended to increase for seed yield and Zn concentration with the addition of genotypes belonging to the same families in the prediction set. Additionally, SSI model showed higher accuracy compared to RKHS for Zn concentration and yield, and RKHS resulted in higher predictions for Fe concentration (**Figure 4.7**). Although the accuracy gain obtained using SSI compared with RKHS is small (<0.05), the prediction of the individuals is provided by, on average, $< 19\%$ of the individuals used to train the model for Fe and Zn concentrations. For yield, the prediction is provided by, on average, $< 73\%$ of the individuals used to train the model (**Supplementary Table 4.4**).

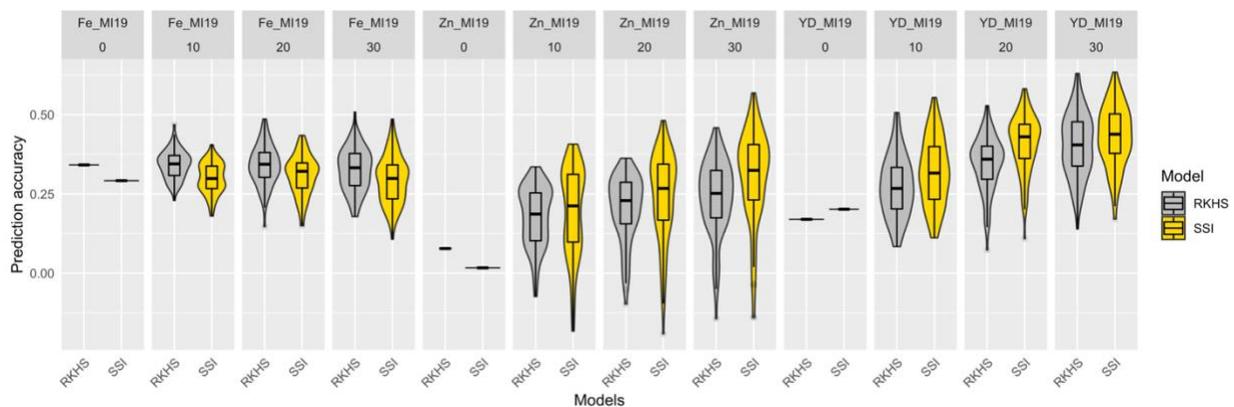


Figure 4.7 Prediction accuracy of Fe-Zn concentration and seed yield (YD) in the 2019 prediction set comprised 82 RILs. Different proportions of the prediction set were included in the models (0%, 10% =8, 20% = 16, 30%=25). The distribution of boxplots represents 100-fold cross-validations.

Discussion

Biofortification is an important goal included in many breeding programs of dry beans. Three assumptions are used to release biofortified cultivars worldwide: *i*) Fe concentration is stable across environments, *ii*) the average Fe concentration in nonbiofortified dry beans is ~ 50 ug/g, *iii*) Fe bioavailability is positively correlated with Fe concentration (Glahn & Noh, 2021). In this

study, we found that none of these assumptions are true in the YBC. Although Fe concentration showed a positive correlation within and between locations in the two growing seasons, those correlations were $< r=0.3$. In both years, beans planted in Nebraska showed a lower Fe concentration (~ 48 ug/g) compared to Michigan (~ 67 ug/g), and the reasons of this difference may be attributed to the alkaline soil (pH = 8) found in the Nebraska location used in this study. Values of pH greater than 7 have been related with iron chlorosis and zinc deficiency in plants (Westfall & Bauder, 2011). Additionally, the Fe and Zn in soil available for plant uptake have been related with the concentration of these microelements in seeds (Katuuramu et al., 2021). The complex genetic architecture of Fe and Zn concentration was described in a meta-QTL analysis in seven populations of dry beans, and candidate genes related with the process to uptake, transport and accumulation of these minerals were identified (Izquierdo et al., 2018). Due the interaction with environmental factors and the complex genetic architecture of Fe and Zn concentration is challenging to get a stable concentration across environments for these microelements.

Our results found the absence of correlation between Fe concentration and in vitro Fe bioavailability. Both years in Michigan showed a non-significant negative correlation between Fe concentration and FeBio, and similar results have been reported in dry beans (Glahn et al., 2020; Katuuramu et al., 2018, 2021), which suggest that increases the Fe concentration is not translated into a higher FeBio. For this absence of relationship, Glahn & Noh (2021) proposed to change the focus of biofortification from Fe concentration to FeBio in dry beans. There are two big advantages of breeding for FeBio over Fe concentration, higher heritability, and delivery of absorbable Fe. We observed that FeBio was very stable across years ($r = 0.93$), with a heritability >0.85 that was 2-fold the heritability of Fe concentration presented in both growing seasons in Michigan.

The high phenotypic variability and heritability of FeBio found in this study suggest that genetic improvement is feasible. However, we observed a strong relationship between seed type and FeBio, which could limit the genetic gain of FeBio in some market classes. The seed types Manteca, Mayocoba and White showed high FeBio values, and we consider that these three market classes should use as priority in biofortification programs. Interestingly, using data reported by Rie et al. 2021, we identified that lighter color seeds from the YBC tend to have lower cooking times. Thicker seed coats were associated with longer cooking times, and darker seeds tend to have thicker seed coats (Bassett et al., 2021) Fast-cooking time has been associated with high FeBio and with more retention of nutrients in dry beans (Wiesinger et al., 2016), which means that the

use of Manteca, Mayocoba and White seed types have the potential to yield high FeBio and nutrients in fast-cooking genotypes, traits that are appealing for nutrition and customer acceptance.

Genome-wide association (GWA)

A strong environment effect was observed for yield and Fe and Zn concentrations, and no consistent associations were identified for these traits. The complex architecture of these traits and the G x E interaction has yielded hundreds of marker-trait associations across locations and populations (Izquierdo et al., 2018, Izquierdo et al., under review), and the usage of strategies as marker-assisted selection or GWAS-assisted genomic prediction does not appear promising for these traits.

FeBio showed large variability as well but this variability across years and individuals was mostly controlled by genetics. The binomial distribution of FeBio is probably because the trait is controlled by major QTL. The SNP in chromosome 7 at 29.2 Mb was associated with light-colored seed, hilum ring and corona and resistance to darkening in the YBC (Sadohara et al., 2021). This SNP is proximal to the P gene (Phvul.007G171333) located at 28.8 Mb, a transcription factor required for flavonoid biosynthesis, which has the dominant allele P necessary for seed coat color expression, and in recessive genotypes, produces white seed coat (McClellan et al., 2018). This region in chromosome 7 was associated with FeBio in the Andean Diversity panel (Katuuramu et al., 2018), and dark-colored bean seeds have been shown to have low FeBio (Wiesinger et al., 2018).

This result gives more evidence of the relationship between seed color and FeBio. There is a correlation between color seed and FeBio, and although unclear, the relationship could be explained for several effects. First, darker seed types tend to have thicker seed coats, and although ~30% of minerals are in the seed coat (Glahn et al., 2016), this tissue shows a high level of antinutrients that can chelate Fe. Besides, in data reported by Sadohara et al., (2021), cooking time in darker seeds tend to be longer than in lighter color seeds, this trait could be related to the thicker seed coat observed in darker seeds. Considering that *i*) the SNP effect in chromosome 7 was close to the mean of FeBio (50 % of control) in both years (**Table 4.1**), *ii*) the association was presented across years, and *iii*) the association with FeBio was present in other genetic background as the Andean Diversity panel, this region can be related with a major QTL for FeBio.

Genomic prediction – training set

In this study, we consider RKHS and SSI models to assess the accuracy of genomic prediction in the YBC. We observed similar results among models for all traits, and the prediction variation was related to the heritability of each trait. Although both models gave similar results in the YBC, the individuals that support the prediction of each line in the testing set are different. RKHS models used in each partition all the genotypes in the training subset, while SSI selected a particular set of support points as the optimal training set for each genotype in the testing set. The average number of supporting points selected by the SSI model was related to the complexity of each trait, ranging from 38% (FeBio) to 71% (yield). In a multigeneration wheat dataset across 8 years, SSI presented higher prediction accuracy compared to GBLUP (Lopez-Cruz et al., 2022), and as demonstrated by Lopez-Cruz & De Los Campos (2021), SSI tends to give higher accuracies compared to GBLUP in large datasets. However, the increase in prediction accuracy of SSI compared with GBLUP tends to be modest (<0.05) and is highly influenced by the diversity and training size. The YBC is a diverse population that can take advantage of SSI, but the training set was comprised by < 185 compared to the thousands of individuals used to train the models by Lopez-Cruz & De Los Campos (2021) and Lopez-Cruz et al. (2022).

To assess the implementation of genomic prediction using the YBC as a training population, 82 breeding lines were used as the prediction dataset. In chapter 3, we report evidence that updating the models with genotypes related to the prediction set tends to increase accuracy. By adding different percentages of each RIL family, prediction accuracy increased for seed yield and Zn, and SSI outperformed RKHS on these traits. The better performance of SSI compared to RKHS in this scenario could be the result of differences in allele frequencies and LD patterns in the families to be predicted compared to the YBC. Differences in allele frequencies and LD may lead to suboptimal estimation of breeding values using the all set of individuals to train the model (Lopez-Cruz et al., 2022).

The polygenic inheritance of the traits involved in this study is reflected in the continuous variation presented in both locations and years. However, the use of consistent QTL associated with FeBio as fixed variables in RKHS model improve accuracy on average by 0.1 (2018) and 0.07 (2019) compared with the model without fixed effects. Similar results were observed in a dry bean black breeding panel on traits with high heritability as color retention of canned beans and seed weight (chapter 3). It is noteworthy that the positive effect in prediction ability of the fixed

markers in this study are potentially related with the high diversity in seed colors found in the YBC, and their positive effect could be reduced in germplasm with less diverse seed types. Nevertheless, accuracy for FeBio was high (> 0.65) using RKHS and SSI models without the inclusion of fixed markers, which suggest that it is possible a reliable prediction of this trait in dry beans. Interestingly, fixed markers negatively affected the accuracy of SSI and reduced the average number of supporting points for predictions. The reduced number of supporting points may affect the estimation of the fixed effects, reducing prediction ability using this approach.

Conclusion

This study provides evidence regarding the need to move from the traditional biofortification approach of increasing Fe concentration to produce dry bean cultivars with more bioavailable Fe. The absence of correlation between Fe concentration and Fe bioavailability observed in this study is supported by studies deployed in Africa and US (Katuramu et al., 2018, 2021), demonstrating that a reallocation of resources may be beneficial in breeding programs focusing on biofortification. We found a strong relationship between Fe bioavailability and seed coat color, and identified that lighter seed color such as Manteca, Mayocoba and White tend to have higher Fe bioavailability and must be priority to enhance the delivery of Fe for human consumption. Despite the importance of Fe bioavailability, the measurement of this trait may be inaccessible for most breeding programs. We found that genomic prediction has high accuracy for Fe bioavailability and can be employed to reduce this costly and time-consuming measurement.

BIBLIOGRAPHY

- Bassett, A., Hooper, S., & Cichy, K. (2021). Genetic variability of cooking time in dry beans (*Phaseolus vulgaris* L.) related to seed coat thickness and the cotyledon cell wall. *Food Research International*, 141(November 2020). <https://doi.org/10.1016/j.foodres.2020.109886>
- Beebe, S. (2020). Biofortification of Common Bean for Higher Iron Concentration. *Frontiers in Sustainable Food Systems*, 4(November), 1–6. <https://doi.org/10.3389/fsufs.2020.573449>
- Bernardo, R. (2020). Breeding for Quantitative Traits in Plants. In *Stemma Press (Issue Third Edition)*. <https://doi.org/10.2135/cropsci2003.1578>
- Birachi, E. A., Sperling, L., Kadege, E., Mdachi, M., Upendo, T., Radegunda, K., Mutua, M., Mbiu, J., Raya, N., & Ndunguru, A. (2020). Analysis of the Yellow Bean Corridor in Tanzania.
- Blair, M. W., Astudillo, C., Rengifo, J., Beebe, S. E., & Graham, R. (2011). QTL analyses for seed iron and zinc concentrations in an intra-genepool population of Andean common beans (*Phaseolus vulgaris* L.). *TAG. Theoretical and Applied Genetics. Theoretische Und Angewandte Genetik*, 122(3), 511–521. <https://doi.org/10.1007/s00122-010-1465-8>
- Blair, M. W., & Izquierdo, P. (2012). Use of the advanced backcross-QTL method to transfer seed mineral accumulation nutrition traits from wild to Andean cultivated common beans. *TAG. Theoretical and Applied Genetics. Theoretische Und Angewandte Genetik*, 125(5), 1015–1031. <https://doi.org/10.1007/s00122-012-1891-x>
- Blair, M. W., Izquierdo, P., Astudillo, C., & Grusak, M. a. (2013). A legume biofortification quandary: variability and genetic control of seed coat micronutrient accumulation in common beans. *Frontiers in Plant Science*, 4(July), 275–289. <https://doi.org/10.3389/fpls.2013.00275>
- Browning, B. L., Zhou, Y., & Browning, S. R. (2018). A One-Penny Imputed Genome from Next-Generation Reference Panels. *American Journal of Human Genetics*, 103(3), 338–348. <https://doi.org/10.1016/j.ajhg.2018.07.015>
- Cichy, K. A., Caldas, G. V., Snapp, S. S., & Blair, M. W. (2009). QTL analysis of seed iron, zinc, and phosphorus levels in an andean bean population. *Crop Science*, 49(5), 1742–1750. <https://doi.org/10.2135/cropsci2008.10.0605>
- Cichy, K. A., Wiesinger, J. A., & Mendoza, F. A. (2015). Genetic diversity and genome-wide association analysis of cooking time in dry bean (*Phaseolus vulgaris* L.). *Theoretical and Applied Genetics*, 128(8), 1555–1567. <https://doi.org/10.1007/s00122-015-2531-z>
- Cichy, K., Chiu, C., Isaacs, K., & Glahn, R. (2022). Dry Bean Biofortification with Iron and Zinc. In *Biofortification of Staple Crops*. https://doi.org/10.1007/978-981-16-3280-8_10
- de los Campos, G., Vazquez, A. I., Fernando, R., Klimentidis, Y. C., & Sorensen, D. (2013). Prediction of Complex Human Traits Using the Genomic Best Linear Unbiased Predictor. *PLoS Genetics*, 9(7). <https://doi.org/10.1371/journal.pgen.1003608>

- Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. a, Kawamoto, K., Buckler, E. S., & Mitchell, S. E. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PloS One*, 6(5), e19379. <https://doi.org/10.1371/journal.pone.0019379>
- Glahn, R. P., Lee, O. A., Yeung, A., Goldman, M. I., & Miller, D. D. (1998). Caco-2 cell ferritin formation predicts nonradiolabeled food iron availability in an in vitro digestion/Caco-2 cell culture model. *Journal of Nutrition*, 128(9), 1555–1561. <https://doi.org/10.1093/jn/128.9.1555>
- Glahn, R. P., & Noh, H. (2021). Redefining Bean Iron Biofortification: A Review of the Evidence for Moving to a High Fe Bioavailability Approach. *Frontiers in Sustainable Food Systems*, 5(July), 1–9. <https://doi.org/10.3389/fsufs.2021.682130>
- Glahn, R. P., Tako, E., Cichy, K., & Wiesinger, J. (2016). The cotyledon cell wall and intracellular matrix are factors that limit iron bioavailability of the common bean (*Phaseolus vulgaris*). *Food and Function*, 7(7), 3193–3200. <https://doi.org/10.1039/c6fo00490c>
- Glahn, R. P., Wiesinger, J. A., & Lung’Aho, M. G. (2020). Iron concentrations in biofortified beans and nonbiofortified marketplace varieties in East Africa are similar. *Journal of Nutrition*, 150(11), 3013–3023. <https://doi.org/10.1093/jn/nxaa193>
- Isidro y Sánchez, J., & Akdemir, D. (2021). Training Set Optimization for Sparse Phenotyping in Genomic Selection: A Conceptual Overview. *Frontiers in Plant Science*, 12(September), 1–14. <https://doi.org/10.3389/fpls.2021.715910>
- Izquierdo, P., Astudillo, C., Blair, M. W., Iqbal, A. M., Raatz, B., & Cichy, K. A. (2018). Meta-QTL analysis of seed iron and zinc concentration and content in common bean (*Phaseolus vulgaris* L.). *Theoretical and Applied Genetics*, 131(8), 1645–1658. <https://doi.org/10.1007/s00122-018-3104-8>
- Katuuramu, D. N., Hart, J. P., Porch, T. G., Grusak, M. A., Glahn, R. P., & Cichy, K. A. (2018). Genome-wide association analysis of nutritional composition-related traits and iron bioavailability in cooked dry beans (*Phaseolus vulgaris* L.). *Molecular Breeding*, 38(4), 44. <https://doi.org/10.1007/s11032-018-0798-x>
- Katuuramu, D. N., Wiesinger, J. A., Luyima, G. B., Nkalubo, S. T., Glahn, R. P., & Cichy, K. A. (2021). Investigation of Genotype by Environment Interactions for Seed Zinc and Iron Concentration and Iron Bioavailability in Common Bean. *Frontiers in Plant Science*, 12(May), 1–10. <https://doi.org/10.3389/fpls.2021.670965>
- Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, 9(4), 357–359. <https://doi.org/10.1038/nmeth.1923>
- Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*, 27(21), 2987–2993. <https://doi.org/10.1093/bioinformatics/btr509>

- Lobaton, J. D., Miller, T., Gil, J., Ariza, D., Hoz, J. F., Soler, A., Beebe, S., Duitama, J., Gepts, P., & Raatz, B. (2018). Resequencing of Common Bean Identifies Regions of Inter-Gene Pool Introgression and Provides Comprehensive Resources for Molecular Breeding. *The Plant Genome*, 11(2), 170068. <https://doi.org/10.3835/plantgenome2017.08.0068>
- Lopez-Cruz, M., Beyene, Y., Gowda, M., Crossa, J., Pérez-Rodríguez, P., & de los Campos, G. (2021). Multi-generation genomic prediction of maize yield using parametric and non-parametric sparse selection indices. *Heredity*, 127(5), 423–432. <https://doi.org/10.1038/s41437-021-00474-1>
- Lopez-Cruz, M., & De Los Campos, G. (2021). Optimal breeding-value prediction using a sparse selection index. <https://doi.org/10.1093/genetics/iyab030>
- Lopez-Cruz, M., Dreisigacker, S., Crespo-Herrera, L., Bentley, A. R., Singh, R., Poland, J., Shrestha, S., Huerta-Espino, J., Govindan, V., Juliana, P., Mondal, S., Pérez-Rodríguez, P., & Crossa, J. (2022). Sparse kernel models provide optimization of training set design for genomic prediction in multiyear wheat breeding data. *Plant Genome*, July, 1–15. <https://doi.org/10.1002/tpg2.20254>
- Lopez-Cruz, M., Olson, E., Rovere, G., Crossa, J., Dreisigacker, S., Mondal, S., Singh, R., & Campos, G. de los. (2020). Regularized selection indices for breeding value prediction using hyper-spectral image data. *Scientific Reports*, 10(1), 1–12. <https://doi.org/10.1038/s41598-020-65011-2>
- Lorenz, A. J., & Smith, K. P. (2015). Adding genetically distant individuals to training populations reduces genomic prediction accuracy in Barley. *Crop Science*, 55(6), 2657–2667. <https://doi.org/10.2135/cropsci2014.12.0827>
- McClellan, P. E., Bett, K. E., Stonehouse, R., Lee, R., Pflieger, S., Moghaddam, S. M., Geffroy, V., Miklas, P., & Mamidi, S. (2018). White seed color in common bean (*Phaseolus vulgaris*) results from convergent evolution in the P (pigment) gene. *New Phytologist*, 219(3), 1112–1123. <https://doi.org/10.1111/nph.15259>
- Meuwissen, T. H. E., Hayes, B. J., & Goddard, M. E. (2001). Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps.
- Perea, C., De La Hoz, J. F., Felipe Cruz, D., Lobaton, J. D., Izquierdo, P., Quintero, J. C., Raatz, B., & Duitama, J. (2016). Bioinformatic analysis of genotype by sequencing (GBS) data with NGSEP. *BMC Genomics*, 17(5), 498. <https://doi.org/10.1186/s12864-016-2827-7>
- Rodríguez-Álvarez, M. X., Boer, M. P., van Eeuwijk, F. A., & Eilers, P. H. C. (2018). Correcting for spatial heterogeneity in plant breeding experiments with P-splines. *Spatial Statistics*. <https://doi.org/10.1016/j.spasta.2017.10.003>
- Sadohara, R., Izquierdo, P., Couto Alves, F., Porch, T. G., Beaver, J. S., Urrea, C. A., & Cichy, K. A. (2022). The *Phaseolus vulgaris* L. Yellow Bean Collection: Genetic diversity and

- characterization for cooking time (Under review). *Genetic Resources and Crop Evolution*, 0. <https://doi.org/10.1007/s10722-021-01323-0>
- Sadohara, R., Morris, D., Long, Y., Cichy, K., Izquierdo, P., & Urrea, C. A. (2021). Seed coat color genetics and genotype \times environment effects in yellow beans via machine-learning and genome-wide association. September, 1–15. <https://doi.org/10.1002/tpg2.20173>
- Schmutz, J., McClean, P. E., Mamidi, S., Wu, G. A., Cannon, S. B., Grimwood, J., Jenkins, J., Shu, S., Song, Q., Chavarro, C., Torres-Torres, M., Geffroy, V., Moghaddam, S. M., Gao, D., Abernathy, B., Barry, K., Blair, M., Brick, M. a, Chovatia, M., ... Jackson, S. a. (2014). A reference genome for common bean and genome-wide analysis of dual domestications. *Nature Genetics*, 46(November 2013), 707–713. <https://doi.org/10.1038/ng.3008>
- Tako, E., Beebe, S. E., Reed, S., Hart, J. J., & Glahn, R. P. (2014). Polyphenolic compounds appear to limit the nutritional benefit of biofortified higher iron black bean (*Phaseolus vulgaris* L.). 1–9.
- Tako, E., Reed, S., Anandaraman, A., Beebe, S. E., Hart, J. J., & Glahn, R. P. (2015). Studies of cream seeded carioca beans (*Phaseolus vulgaris* L.) from a Rwandan efficacy trial: In vitro and in vivo screening tools reflect human studies and predict beneficial results from iron Biofortified beans. *PLoS ONE*, 10(9), 1–15. <https://doi.org/10.1371/journal.pone.0138479>
- Tello, D., Gil, J., Loaiza, C. D., Riascos, J. J., Cardozo, N., & Duitama, J. (2019). NGSEP3: Accurate variant calling across species and sequencing protocols. *Bioinformatics*, 35(22), 4716–4723. <https://doi.org/10.1093/bioinformatics/btz275>
- Uebersax, M. A., Cichy, K. A., Gomez, F. E., Poch, T. G., Heitholt, J., Osorno, J. M., Kamfwa, K., Snapp, S. S., & Bales, S. (2022). Dry beans (*Phaseolus vulgaris* L.) as a vital component of sustainable agriculture and food security—A review. *Legume Science*, April, 1–13. <https://doi.org/10.1002/leg3.155>
- Van, B.-J. (2022). StatgenGxE: Genotype by Environment (GxE) Analysis. Available from: <https://cran.r-project.org/package=statgenGxE>. <https://doi.org/10.3389/fphys.2013.00044>>.One
- Voss-Fels, K. P., Cooper, M., & Hayes, B. J. (2019). Accelerating crop genetic gains with genomic selection. In *Theoretical and Applied Genetics* (Vol. 132, Issue 3, pp. 669–686). Springer Verlag. <https://doi.org/10.1007/s00122-018-3270-8>
- Wang, J., & Zhang, Z. (2021). GAPIT Version 3: Boosting Power and Accuracy for Genomic Association and Prediction. *Genomics, Proteomics and Bioinformatics*, 19(4), 629–640. <https://doi.org/10.1016/j.gpb.2021.08.005>
- Westfall, D., & Bauder, T. (2011). Zinc and Iron Deficiencies. Colorado State University, Extension, 0.545(0), 0–2.

- Wiesinger, J. A., Cichy, K. A., Glahn, R. P., Grusak, M. A., Brick, M. A., Thompson, H. J., & Tako, E. (2016). Demonstrating a Nutritional Advantage to the Fast-Cooking Dry Bean (*Phaseolus vulgaris* L.). *Journal of Agricultural and Food Chemistry*, 64(45), 8592–8603. <https://doi.org/10.1021/acs.jafc.6b03100>
- Wiesinger, J. A., Cichy, K. A., Tako, E., & Glahn, R. P. (2018). The fast cooking and enhanced iron bioavailability properties of the manteca yellow bean (*Phaseolus vulgaris* L.). *Nutrients*, 10(11). <https://doi.org/10.3390/nu10111609>
- Wiesinger, J. A., Glahn, R. P., Cichy, K. A., Kolba, N., Hart, J. J., & Tako, E. (2019). An in vivo (*Gallus gallus*) feeding trial demonstrating the enhanced iron bioavailability properties of the fast cooking manteca yellow bean (*Phaseolus vulgaris* L.). *Nutrients*, 11(8). <https://doi.org/10.3390/nu11081768>
- Willett, W., Rockström, J., Loken, B., Springmann, M., Lang, T., Vermeulen, S., Garnett, T., Tilman, D., DeClerck, F., Wood, A., Jonell, M., Clark, M., Gordon, L. J., Fanzo, J., Hawkes, C., Zurayk, R., Rivera, J. A., De Vries, W., Majele Sibanda, L., ... Murray, C. J. L. (2019). Food in the Anthropocene: the EAT–Lancet Commission on healthy diets from sustainable food systems. *The Lancet*, 393(10170), 447–492. [https://doi.org/10.1016/S0140-6736\(18\)31788-4](https://doi.org/10.1016/S0140-6736(18)31788-4)

APPENDIX

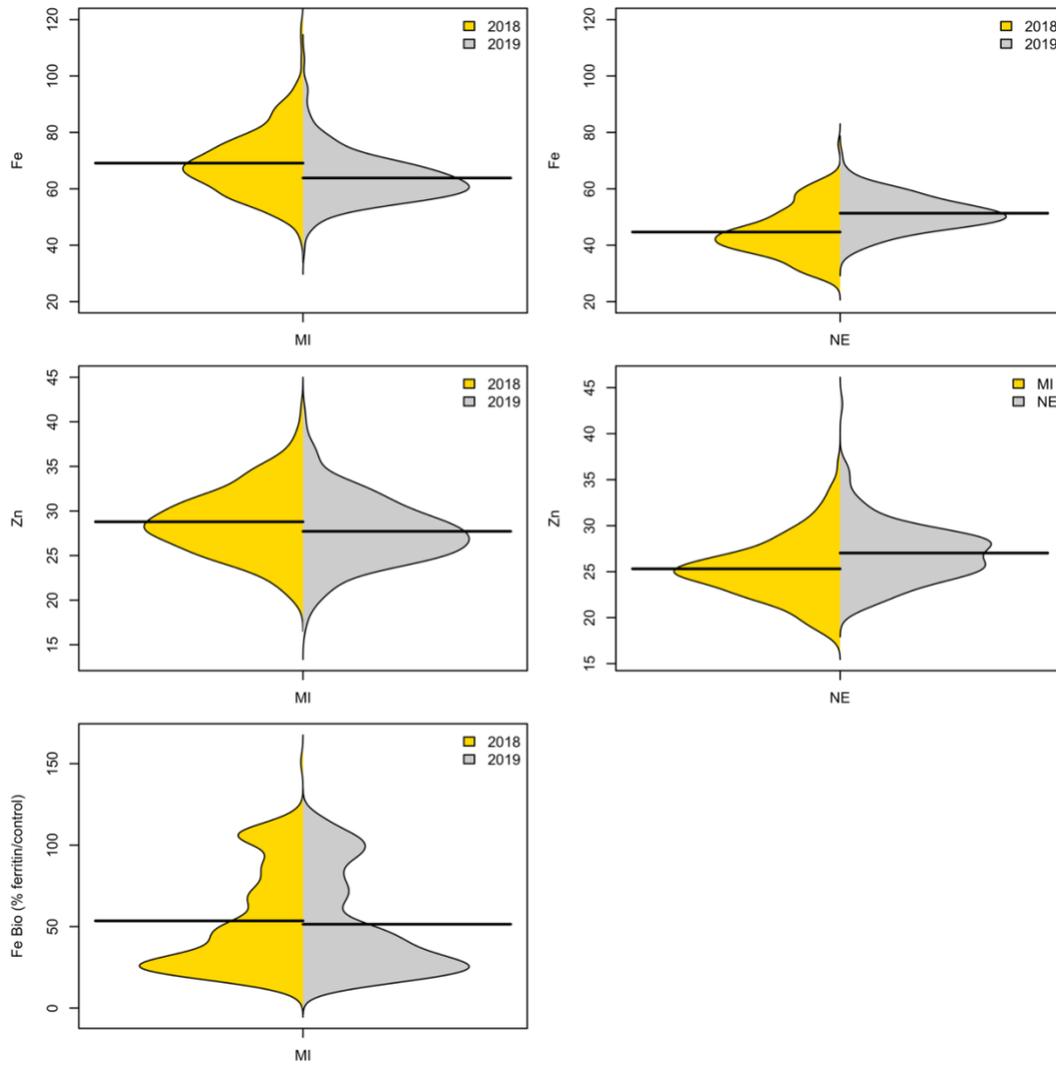


Figure S4.1 Bean plots of Fe, Zn in MI and Ne, and Fe bioavailability in MI. The trials were carried out in 2018 and 2019 in MI and NE. Micronutrients traits were collected in one field replication, and two technical replications were done per sample. The mean of the two technical replications were used as measure for each environment and year.

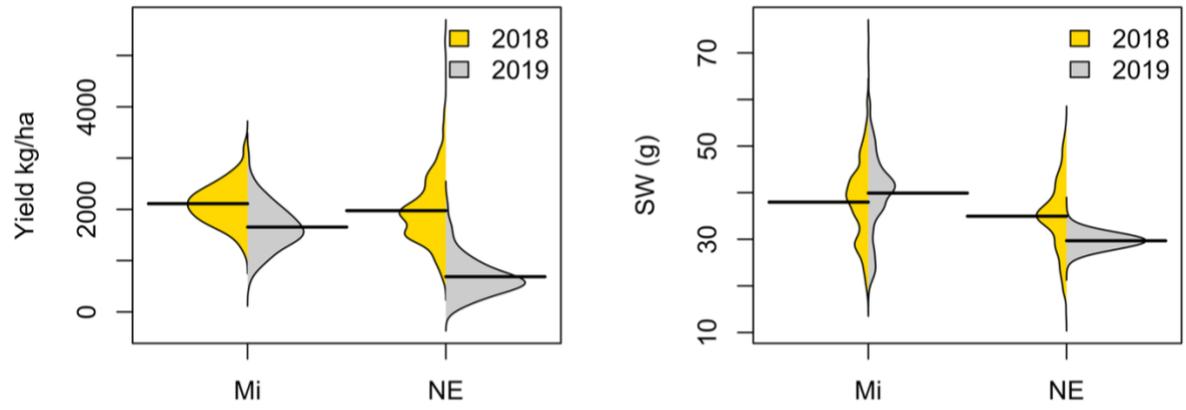


Figure S4.2 Bean plots of yield (YD) and seed weight (SW) of the yellow panel. The trials were carried out in 2018 and 2019 in MI and NE. Best linear unbiased estimator (BLUE) were obtained for each trait and trial, adjusting for spatial effects in the field.

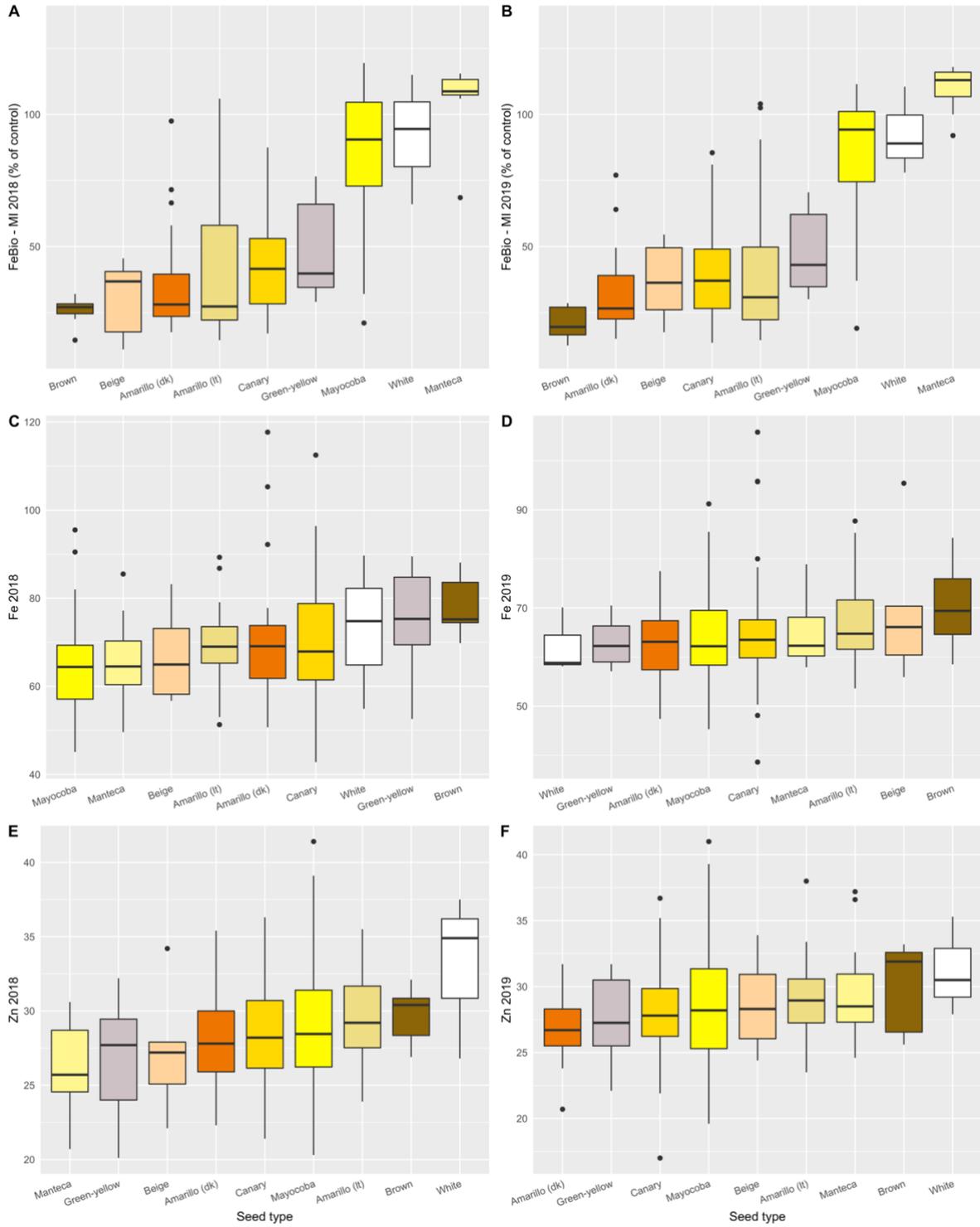


Figure S4.3 Boxplots of Fe-Zn concentration and Fe bioavailability (FeBio) in the nine major seed types in the Yellow Bean Collection (YBC) in MI.

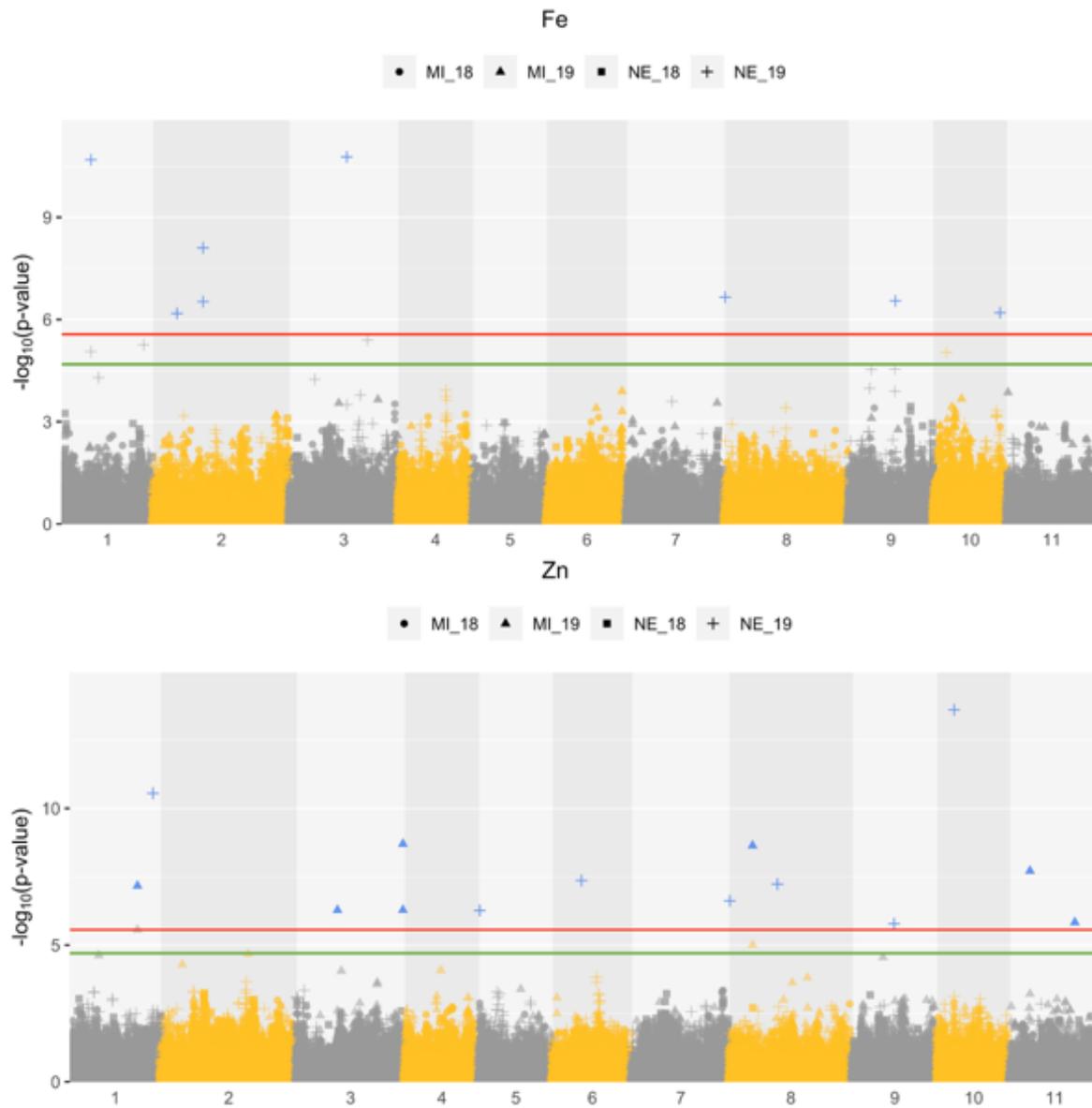


Figure S4.4 Genome-wide association analysis of Fe and Zn. The red and green lines indicate the Bonferroni and FDR threshold at $\alpha = 0.05$, respectively.

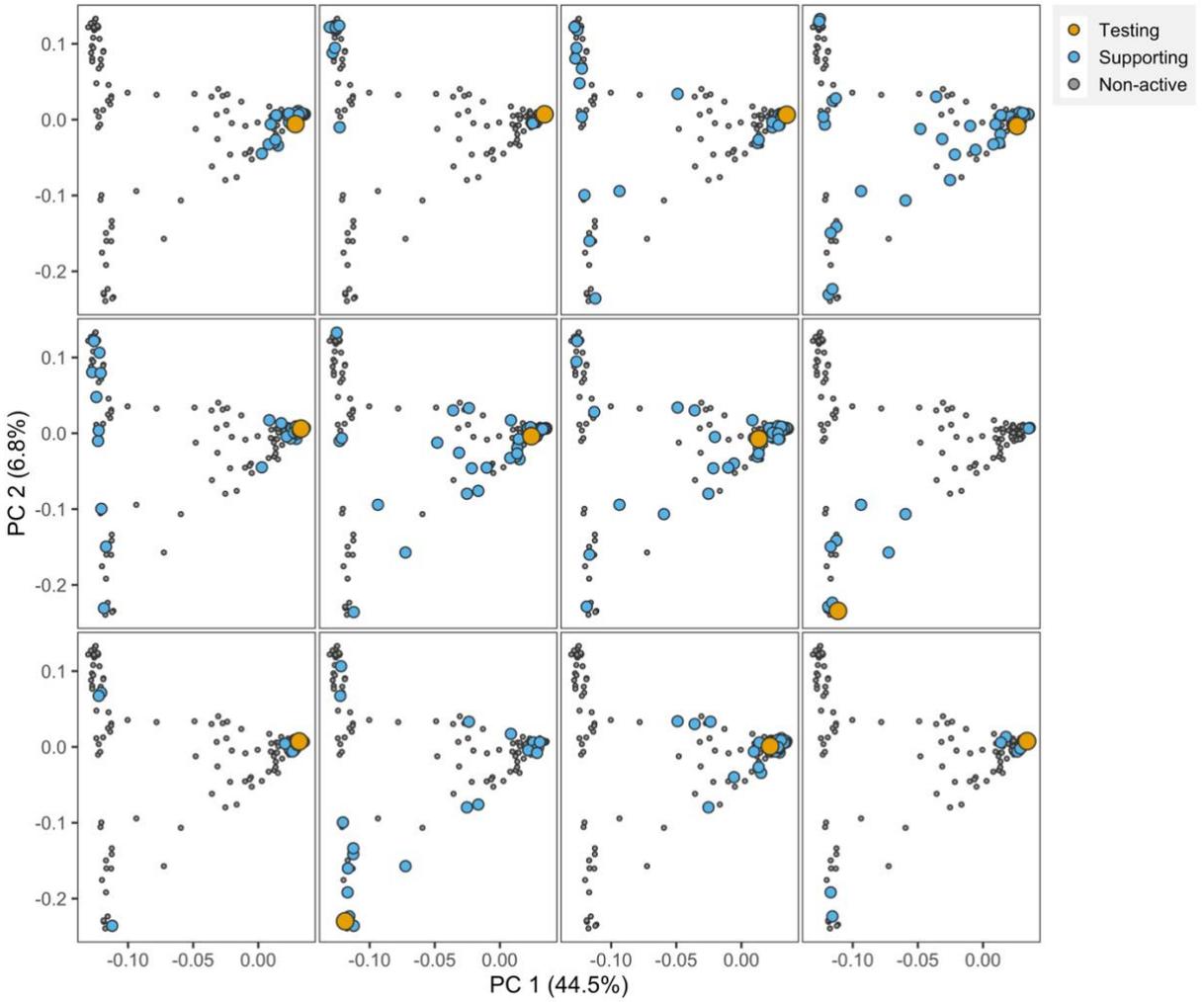


Figure S4.5 Network of genotypes in the training data set that contribute to the prediction for FeBio in 2019. The axes represent the first two principal components coordinates for prediction points (yellow) and the corresponding support points (blue). Gray points represent genotypes that did not contribute to the prediction.

Table S4.1 Analysis of variance.

Source of variation	R2				
	Fe	Zn	FeBio	Yield	SW
Year	0.0	0.0	0.2**	39.5***	0.0**
Location	62.5***	12.4***	-	8.4***	28.0***
Year:location	6.4***	5.5***	-	23.2***	9.1***
Genotype	5.2***	36.7***	93.0***	4.8***	41.6***
Genotype:year	0.0	0.0	-	1.9*	0.0
Genotype:location	3.3*	0.0	-	9.0***	5.9**

Signif. codes: 0 ***, 0.001**, 0.01 *.

Table S4.2 QTL identified by genome-wide association analysis using BLINK and/or FarmCPU models with a Bonferroni correction.

Trait	Chr	Position	P.value	maf	Model	Location	Year
Fe	1	7,292,184	2.06E-11	0.42	FarmCPU	NE	2019
	2	3,540,974	6.63E-07	0.22	FarmCPU	NE	2019
	2	25,542,798	7.87E-09	0.11	Blink-FarmCPU	NE	2019
	3	36,517,097	1.69E-11	0.09	Blink	NE	2019
	8	84,336	2.21E-07	0.11	Blink	NE	2019
	9	25,301,211	2.82E-07	0.18	FarmCPU	NE	2019
	10	43,391,440	6.34E-07	0.15	FarmCPU	NE	2019
FeBC	1	51,031,345	8.11E-09	0.10	Blink-FarmCPU	MI	2018
	3	2,649,825	5.91E-09	0.24	Blink	MI	2019
	3	4,512,266	4.15E-08	0.12	FarmCPU	MI	2019
	4	46,268,498	3.21E-10	0.07	Blink	MI	2019
	7	29,169,848	1.31E-12	0.18	Blink-FarmCPU	MI	2018
	7	29,169,848	3.43E-19	0.18	Blink-FarmCPU	MI	2019
	8	60,316,632	4.24E-08	0.12	Blink	MI	2019
	9	37,951,003	4.96E-07	0.35	FarmCPU	MI	2019
	10	5,414,443	8.93E-08	0.16	Blink	MI	2018
	10	41,221,348	3.00E-08	0.35	Blink	MI	2019
	10	42,422,135	5.90E-11	0.26	Blink-FarmCPU	MI	2018
	10	42,422,135	2.67E-17	0.26	FarmCPU	MI	2019
	10	42,791,842	4.22E-07	0.09	Blink	MI	2019
	10	42,912,633	2.26E-08	0.08	FarmCPU	MI	2019
	11	3,014,294	8.23E-08	0.08	Blink	MI	2018
	11	3,626,904	4.75E-08	0.28	Blink-FarmCPU	MI	2018
	11	3,626,904	6.77E-08	0.28	FarmCPU	MI	2019
	11	4,761,183	6.78E-09	0.19	FarmCPU	MI	2019
11	49,383,743	2.37E-08	0.20	FarmCPU	MI	2019	
11	53,419,997	2.38E-06	0.28	FarmCPU	MI	2019	
ZN	1	47,810,686	6.72E-08	0.13	Blink	MI	2019
	1	50,568,067	2.80E-11	0.17	Blink	NE	2019
	3	19,332,217	5.21E-07	0.30	FarmCPU	MI	2019
	3	53,059,251	1.98E-09	0.16	Blink-FarmCPU	MI	2019
	5	69,261	5.40E-07	0.20	Blink	NE	2019
	6	20,009,852	4.34E-08	0.21	Blink	NE	2019
	8	84,336	2.39E-07	0.11	Blink	NE	2019
	8	4,308,119	2.28E-09	0.32	Blink	MI	2019
	8	22,052,184	5.86E-08	0.15	Blink	NE	2019
	9	21,789,929	1.65E-06	0.27	Blink	NE	2019
	10	4,482,939	2.48E-14	0.10	Blink	NE	2019
	11	2,463,522	1.92E-08	0.29	FarmCPU	MI	2019
	11	47,634,294	1.46E-06	0.40	Blink	MI	2019
YD	3	44,116,362	3.11E-08	0.07	FarmCPU	MI	2019
	3	51,442,534	1.53E-06	0.12	FarmCPU	MI	2018
	3	52,993,055	9.97E-07	0.27	Blink-FarmCPU	MI	2018
	4	13,916,133	6.28E-10	0.33	Blink-FarmCPU	MI	2019
	4	43,890,873	4.50E-07	0.11	Blink	NE	2018
	5	914,931	2.60E-08	0.39	Blink-FarmCPU	MI	2018
	5	4,497,912	8.81E-10	0.30	FarmCPU	NE	2018
	7	35,814,640	7.75E-08	0.37	Blink	MI	2019
	8	60,779,846	2.12E-06	0.16	FarmCPU	MI	2018
	9	14,724,912	1.82E-07	0.09	FarmCPU	NE	2018
	11	3,626,904	4.54E-07	0.39	Blink	MI	2019

Table S4.3 Prediction accuracy for micronutrients density and seed yield in the YBC using RKHS and sparse selection indices (SSI).

Trait	Location	Year	h^2	ntst	ntrn	Model	nsup	Accuracy (sd)
FeBio	MI	2018	0.85	78	185	RKHS	185	0.656(0.06)
						RKHS-QTL	185	0.756(0.04)
		SSI	78	0.654(0.06)				
		SSI-QTL	57	0.603(0.07)				
	2019	0.91	78	185	RKHS	185	0.703(0.06)	
					RKHS-QTL	185	0.774(0.04)	
		SSI	63	0.702(0.06)				
		SSI-QTL	41	0.617(0.08)				
Fe	MI	2018	0.4	78	185	RKHS	185	0.336(0.09)
						SSI	74	0.258(0.13)
		2019	0.42	68	161	RKHS	161	0.229(0.08)
						SSI	54	0.189(0.14)
	NE	2018	0.05	65	154	RKHS	154	0.111(0.10)
						SSI	113	0.129(0.10)
		2019	0.79	63	147	RKHS	147	0.459(0.08)
						SSI	88	0.418(0.10)
Zn	MI	2018	0.49	78	185	RKHS	185	0.405(0.07)
						SSI	100	0.395(0.07)
		2019	0.67	68	161	RKHS	161	0.356(0.08)
						SSI	55	0.354(0.09)
	NE	2018	0.2	65	154	RKHS	154	0.199(0.10)
						SSI	116	0.142(0.12)
		2019	0.78	63	147	RKHS	147	0.517(0.07)
						SSI	86	0.493(0.09)
YD	MI	2018	0.63	78	185	RKHS	185	0.336(0.09)
						SSI	99	0.327(0.09)
		2019	0.72	78	182	RKHS	182	0.647(0.06)
						SSI	123	0.648(0.06)
	NE	2018	0.65	66	157	RKHS	157	0.514(0.09)
						SSI	105	0.508(0.09)
		2019	0.59	65	152	RKHS	152	0.658(0.08)
						SSI	145	0.654(0.08)

Accuracy is reported as the average of 100 partitions using the training set (70%:30%).

Table S4.4 Prediction accuracy for Fe-Zn concentration and seed yield in the prediction set of 82 yellow RILs using RKHS and sparse selection indices (SSI).

Trait	Location	Year	h^2	ntst	ntrn	%*	Model	nsup	Accuracy (sd)				
Fe	Mi	2019	0.42	74	229	0	RKHS	229	0.34				
						0	SSI	40	0.29				
				63	240	10	RKHS	240	0.34(0.04)				
						10	SSI	44	0.30(0.05)				
				57	246	20	RKHS	246	0.34(0.06)				
						20	SSI	51	0.31(0.06)				
				50	253	30	RKHS	253	0.33(0.07)				
						30	SSI	49	0.29(0.08)				
				Zn	MI	2019	0.67	74	229	0	RKHS	229	0.08
										0	SSI	35	0.02
63	240	10	RKHS					240	0.18(0.09)				
		10	SSI					40	0.19(0.14)				
57	246	20	RKHS					246	0.20(0.10)				
		20	SSI					45	0.24(0.14)				
50	253	30	RKHS					253	0.24(0.12)				
		30	SSI					44	0.31(0.14)				
YD	MI	2019	0.72					82	260	0	RKHS	260	0.17
										0	SSI	216	0.2
				70	272	10	RKHS	272	0.27(0.10)				
						10	SSI	212	0.32(0.11)				
				64	278	20	RKHS	278	0.35(0.09)				
						20	SSI	157	0.41(0.09)				
55	287	30	RKHS	287	0.40(0.10)								
		30	SSI	210	0.43(0.10)								

Accuracy is reported as the average of 100 partitions. * Different proportions of the prediction set were included in the models (0%, 10%, 20%, 30%).

CONCLUSIONS

The genetic architecture of quantitative traits is controlled by many loci across the genome. Although genetic gains have been achieved through traditional breeding, understanding the genetics of complex traits can lead to further improvement of these traits. Therefore, this dissertation has focused on the implementation of quantitative genetics approaches to increase the genetic gain of seed yield and end-use quality traits.

Chapter 1 unraveled the genetic complexity of Fe and Zn seed levels. The positive correlation between Fe and Zn seed levels across different genetic backgrounds suggests that the accumulation of these microelements is controlled by a similar movement of Fe and Zn through the plant, ultimately to the seed. Previous studies assessing the human nutritional outcomes associated with biofortification have reported that high Fe concentration is related to higher Fe uptake inhibitors, limiting Fe absorption. However, the interaction of high Fe beans with other components of the diet can enhance the delivery of Fe. The meta-QTL and candidate genes identified can be used to develop molecular markers to increase Fe and Zn levels in dry beans.

Chapter 2 describes a genome-wide landscape of genomic regions identified over the last two decades as associated with seed yield components. This study facilitates the comparison of QTL and GWA associations across populations of dry beans and the integration with related species through comparative genomics. Several meta-QTLs were found by associations reported under drought and sufficient water conditions, which supports the premise that genetic gain for seed yield is not exclusive between these water supplies. Although the triangle of seed yield components involving seed size, seed number per pod, and pods per plant are crucial for seed yield, the partitioning efficiency into seeds seems to be the most important trait to increase the genetic gain for seed yield under drought and sufficient water conditions. The combination of QTL and GWA analyses used in this study increased the resolution to detect candidate genes and identified that selection pressure for seed yield led to focus on similar genomic regions in the Andean and Middle American gene pools.

Chapter 3 characterizes the genetic architecture and assesses the accuracy of genomic prediction of canning quality traits in black beans. The prediction accuracy of the models was related to the heritability of each trait, highlighting the importance of high-quality phenotyping to reduce environmental noise. Breeding lines from two different cycles were used, and prediction accuracy within the same cycle was higher because of stronger genetic relationships due to

common parents used within breeding cycles. Genomic prediction models must be updated to be useful in the breeding pipeline, and affordable high-throughput genotyping approaches such as GBS is an ideal option to obtain the genotype of new breeding cycles each year.

Chapter 4 describes the relationship between seed Fe bioavailability and Fe and Zn concentration in yellow beans. The correlation between Fe bioavailability and concentration was negative, showing that increased Fe concentration in seeds is not translated into more absorbable Fe. A change of approach to increase Fe bioavailability instead of Fe concentration can be beneficial in biofortification programs to increase Fe delivery. The heritability of Fe bioavailability was consistently higher than the heritability of Fe and Zn concentration, which was reflected in the higher accuracy of genomic prediction for bioavailability. The high phenotypic variability and heritability of Fe bioavailability found in this study suggest that genetic improvement is feasible. However, there is a strong relationship between seed type and absorbable Fe, which could limit the genetic gain of Fe bioavailability in some market classes. The seed types Manteca, Mayocoba, and white showed high Fe bioavailability values and should use as a priority in biofortification programs.

The studies presented herein provide genetic information to help breeders increase the genetic gain of seed yield, nutritional and canning quality traits in dry beans. Additionally, the findings of using different models and strategies to increase the accuracy of genomic prediction will be useful for breeders to tune up the implementation of genomic prediction.